Biomedical Engineering Application: Disease Diagnosis and Treatment

by

Wei Zhang

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Computer Engineering

Department of Electrical and Computer Engineering

University of Alberta

Abstract

Nowadays, with the rapid development of science and technology, human healthcare has become a hot topic and attracts more and more attention. A lot of researchers work on different technologies to contribute to our healthcare no matter disease diagnosis and prognosis or disease treatment. In this thesis, we first introduce two different automated diseases diagnosis approaches and then design a novel gene delivery system that can help treat genetic diseases.

Artificial intelligence (AI) is a popular research topic now and a lot of researchers are working on it. AI has various successful applications in computer vision, automatic speech recognition, natural language processing, audio recognition, bioinformatics and has been proven to be used in disease diagnosis. Two automated diseases diagnosis approaches are designed for depression and tuberculosis using different machine learning (ML) algorithms, respectively. Depression is one of the most common mental disorders, and rates of depression in individuals continuously increase each year. Traditional diagnosis methods are mostly based on the professional judgment of mental health, which is prone to individual bias. Therefore, it is crucial to design an effective and robust model for automated depression detection. I proposed a multimodal fusion model comprised of text, audio, and video for both depression detection and assessment tasks. For the text modality, a pre-trained sentence embedding algorithm was utilized to extract semantic representation along with bidirectional Long Short-Term Memory

(BiLSTM) to predict depression. We also used principal component analysis (PCA) to reduce the dimensionality of the input feature space and fed it into a support vector machine (SVM) to predict depression based on audio modality. For the video modality, XGBoost was employed to conduct both feature selection and depression detection. The final predictions were given by outputs of different modalities with an ensemble voting algorithm. Experiments on the Distress Analysis Interview Corpus Wizard-of-Oz (DAIC-WOZ) dataset showed our proposed model outperforms the baseline in both depression detection and assessment tasks and has comparable performance with other existing state-of-the-art depression detection methods.

Tuberculosis (TB) is a major public health burden affecting about a quarter of the world's population annually according to the World Health Organization(WHO). Among many control steps, early diagnosis and treatment of TB are critical. Commonly used diagnostic techniques, such as X-ray, TB culture test, TB skin test, and Sputum acid-fast bacillus, have their major limitations. Therefore, new cost-effective diagnostic methods are urgently needed. In this thesis, we used high-resolution liquid chromatography-mass spectrometry (LC-MS) to screen 191 blood samples and discovered kynurenine (Kyn), tryptophan (Trp) and their ratio, Indoleamine 2, 3-dioxygenase (IDO) are excellent TB biomarkers. We employed the logistic regression algorithm to detect pulmonary TB and got excellent performance for classifying health control (HC) vs active tuberculosis (ATB) and latent tuberculosis infection (LTBI) vs ATB. When we used IDO and t-spot to distinguish between nontuberculous lung disease

(NTB) and ATB, the results are always satisfying both on the validation set and external independent cohort.

For the gene delivery system, we synergistically combine non-viral chemical materials, magnetic nanoparticles (MNPs), and physical technique, low-intensity pulsed ultrasound (LIPUS), to achieve efficiently and targeted gene delivery. The MNPs are iron oxide super-paramagnetic nanoparticles, coated with polyethyleneimine (PEI) giving a highly positively charged surface, which is favorable for the binding of genetic materials. Driven by the paramagnetic properties of the MNPs, the application of an external magnetic field increases transfection efficiency while LIPUS stimulation enhances cell viability and permeability. By combining the effect of the external magnetic field and LIPUS, the genetic material (GFP or Cherry Red plasmid) can enter the cells. The flow cytometry results showed that by using just a magnetic field to direct the genetic material, the transfection efficiency on HEK 293 cells that were treated by our MNPs coupled with LIPUS stimulation, increased a lot and was much higher than the positive control (Lipofectamine 2000) and showed less toxicity. Cell viability after transfection was greatly promoted compared to the standard transfection technique.

Preface

This dissertation is submitted for the degree of Doctor of Philosophy at the University of Alberta. This Ph.D. thesis is based on the research performed at the Department of Electrical and Computer Engineering, University of Alberta, under the supervision of Professor Jie Chen between September 2016 and December 2021. Some chapters contain published articles authored or coauthored. All the authors' contributions on the related published articles are listed as follows:

Chapter 2 of this thesis has not yet been published, but we have submitted it as Wei Zhang, Kaining Mao, Jie Chen. "A Multimodal Approach for Detection and Assessment of Depression Using Text, Audio and Video." to Scientific Reports for review. I was responsible for experiments design, conduct, data analysis, and manuscript composition. Dr. Jie. Chen supervised the work, provided valuable guidance, and revised the manuscript.

Chapter 3 of this thesis has not yet been published, but we have prepared as Wei Zhang, Zhenyan Chen, Jie Chen. "Metabolomic Biomarker Selection for Pulmonary Tuberculosis Diagnosis and Prognosis." I was responsible for data collection design, experiments design, data analysis, and manuscript composition. Zhenyan Chen finished the data collection. Dr. Jie. Chen supervised the work, provided valuable guidance, and revised the manuscript.

Chapter 4 of this thesis contains the research published as Wei Zhang, Gaser N.

Abdelrasoul, Oleksandra Savchenko, Abdalla Abdrabou, Zhixiang Wang, Jie Chen. “Ultrasound-assisted magnetic nanoparticle-based gene delivery.” PloS one 15, no. 9 (2020): e0239633. I was responsible for the experiments design, measurement, data analysis, and manuscript composition. Dr. Jie. Chen supervised the work, provided valuable guidance, and revised the manuscript.

Acknowledgments

First and foremost, I would like to express my appreciation to my supervisor Dr. Jie Chen for providing me the valuable opportunity to pursue the Doctor of Philosophy degree at the University of Alberta and perform these projects in his research group. His guidance, patience, critical evaluation of my progress, yet ever-present support are greatly appreciated, without them I would not have completed this thesis.

I would like to express my appreciation to the members of my thesis committee: Dr. Zhixiang Wang and Dr. Gregory Kish for the thorough reading of my thesis and the valuable suggestions to further improve the thesis. I would also like to acknowledge the professors who teach me, including Dr. Bruce Cockburn, Dr. Jie Han, Dr. Venkata Dinavahi, Dr. Scott Dick, Dr. Majid Khabbazian, and Dr. Di Niu. Thank you for helping me build solid background knowledge for my research. I also would like to thank Shiang Qi, Kaining Mao, Xiaoxue Jiang, Yufeng Li, Dr. Gaser N. Abdelrasoul, Dr. Oleksandra Savchenko, and the other group members for their support and help. I also would like to thank Shanghai Public Health Clinical Center for helping collect the data.

Finally, I would like to thank my parents and friends for their unswerving support and continuous encouragement throughout my years of study and throughout my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI Artificial Intelligence

ANOVA Analysis of Variance

ATB Active Tuberculosis

AUC Area Under Curve

AVEC Audio/Visual Emotion Challenge and Workshops

BCE Binary Cross-Entropy

CNN Convolutional Neural Network

DAIC-WOZ Distress Analysis Interview Corpus Wizard-of-Oz

DBN Deep Belief Network

DL Deep Learning

DNN Deep Neural Network

ECG Electrocardiograph

F0 Fundamental Frequency

FCS Fetal Calf Serum

H1 First Harmonics of the Differentiated Glottal Source Spectrum

H2 Second Harmonics of the Differentiated Glottal Source Spectrum

HAM-D Hamilton Depression Rating Scale

HC Healthy Control

HEK Human Embryonic Kidney

HMPDD Harmonic Model and Phase Distortion Deviations

HMPDM Harmonic Model and Phase Distortion Mean

IDO Indoleamine 2, 3-dioxygenase

KNN K-Nearest Neighbors

Kyn Kynurenine

LIPUS Low-Intensity Pulsed Ultrasound

LSTM Long Short-Term Memory

LTBI Latent Tuberculosis Infection

MAE Mean Absolute Error

MCEP Mel Cepstral Coefficient

MDQ Maxima Dispersion Quotient

MEM Minimum Essential Medium

MFCC Mel-Frequency Cepstral Coefficients

MHI Motion History Image

ML Machine Learning

MNP Magnetic Nanoparticle

MSE Mean Squared Error

MTB Mycobacterium Tuberculosis

NAQ Normalized Amplitude Quotient

NTB Nontuberculous Lung Disease

PBS Phosphate Buffered Saline

PCA Principal Component Analysis

PEI Polyethyleneimine

PFA Paraformaldehyde

PSP Parabolic Spectral Parameter

QOQ Quasi-Open Quotient

RMSE Root Mean Squared Error

RNN Recurrent Neural Network

ROC Receiver-Operating Characteristic

SVM Support Vector Machine

TB Tuberculosis

Trp Tryptophan

VUV Voiced/Unvoiced

WHO World Health Organization

1 Introduction

1.1 Artificial Intelligence

AI is a branch of computer science concerned with building intelligent machines capable of performing tasks that typically require human intelligence[1]. Humans, animals, and many machines have different types and degrees of ability to finish the computational part to achieve goals, that is called intelligence[1]. Cause AI is dealing with all aspects of simulating cognitive functions to solve real-world problems and build systems that learn and think like humans[2], it is often referred to as machine intelligence[3] to compare it with human intelligence[4]. Alan Turing (1950) is one of the founders of modern computers and AI. The 'Turing Test' is based on the fact that the intelligent behavior of a computer is the ability to achieve human-level performance in cognitive-related tasks[5]. Since then, this field revolving around the intersection of cognitive science and computer science began to develop rapidly[6]. Five years later, the proof of the concept of AI was first initialized through Logic Theorist, which was a program designed to imitate the problem-solving ability of humans and is considered by many to be the first AI program[7]. The importance of this event cannot be underestimated because it greatly catalyzes future AI research. AI flourished, and computers can now store more information and become faster, cheaper, and more accessible. Many landmark goals of AI have been achieved, such as the success of

AlphaGo. We are now living in the ‘big data’ era, where we can collect a large amount of cumbersome and unable to process information. The application of AI in this area has achieved fruitful results in many industries around us such as technology, banking, marketing, and entertainment.

AI has now aroused great interest of many researchers and flourished largely because of the success of the practical application of ML. ML is a very practical field of AI, aiming to build software that can automatically learn from existing data to acquire knowledge from experience and gradually improve itself to make predictions on new data[8]. It is one of the most rapidly growing research fields and is also the core of AI and data science, at the intersection of computer science and statistics[9]. Over the past 20 years, ML has made tremendous progress, from laboratory curiosity to practical technology for a wide range of commercial uses. Many researchers find that, for most applications, it is much easier to train the system by showing the system examples of the desired input and output behavior than by manually programming be predict the desired response for all possible inputs[9]. The impact of ML is also widely used in computer science and a range of industries[9]. In ML, we usually get a training set and a test set. The training set means the union of the labeled set and the unlabeled set of observations available to the machine learners. In contrast, the test set consists of examples that have never been seen before[10]. According to the nature of the training data, we can divide ML into follows, supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, and transfer learning[11]. Supervised

learning algorithms are trained using the labeled dataset, these datasets are designed to train or ‘supervise’ the algorithms to predict more accurate labels on the test set[11]. Using labeled inputs and outputs, the model can measure its accuracy and learn over time. Supervised learning can be further divided into classification (predicting categorical labels) and regression (predicting continuous labels). Unsupervised learning algorithms are used on data without labels[11]. These algorithms have no ‘right answer’ and the goal is to explore the data and find some structure within. They are mainly used for three tasks, clustering, association, and dimensionality reduction. Semi-supervised learning uses both labeled and unlabeled data for training and has the same applications as supervised learning[11]. It is exactly useful when the cost associated with labeling is too high for a fully labeled training process. Reinforcement learning algorithms use three primary components, the agent, the environment, and the actions to discover which actions produce the greatest rewards through trial and error[11]. The goal of reinforcement learning is to learn the best strategy. It is widely used in robotics, games, and navigation.

DL is a family of ML models based on deep neural networks having a long history. DL is to learn the inherent laws and representation levels of examples based on a large amount of data. The different levels of information obtained in the learning process are very helpful for the interpretation of the text, images, voice, and other kinds of data. DL is a complex ML algorithm based on a large scale of training data. It is very popular today because it has achieved better results than previous related technologies in many

fields especially in speech and image recognition, sometimes even better than human-level performance[2], [10]. Research in this DL attempts to make better representations and learn these representations from a large amount of unlabeled data. Various DL architectures such as DNN, CNN, DBN, and RNN have been applied in computer vision, automatic speech recognition, natural language processing, audio recognition, and bioinformatics where they have achieved the most advanced results. Although DL has been characterized as a buzzword[12], [13], it has shown superior capabilities to other ML algorithms in various tasks[14].

AI is a technology that is rapidly adopted by various industries, mainly focused on improving performance, accuracy, time efficiency, reducing costs, and liberating manpower.

1.2 Artificial Intelligence Application on Disease Diagnosis and Prognosis

AI also has wide applications in medicine with a long history[15]. Since the last century, researchers have been exploring the potential applications of AI technologies in various fields of medicine[15], [16]. The application of AI technology in the field of medicine was first investigated by Gunn in 1976 when he conducted the diagnosis of acute abdominal pain with computer analysis[17]. In the past two decades, there is a great surge in the interest in medical AI, and the potential of AI to leverage meaningful relationships in datasets has been successfully used in many clinical scenarios such as

diagnosis, treatment, prognosis, and predicting[18]. The great challenge that modern medicine faced is how to acquire and apply the experience and knowledge to solve complex clinical problems. The more experience and knowledge we have, the better knowledge-based decisions we can make. These experiences and knowledge are derived from data, the data comes from evidence-based medicine, while the experience comes from the actual results of patients[5].

ML technology has been well studied and applied in the analysis of the medical field, especially in the medical diagnosis of specialized diagnosis problems[19]. The correct diagnosis mostly relies on the clinical data, provided in the form of medical records in specialized hospitals or departments. All that needs to be done is to feed the patients' records with correct label diagnoses into a computer program to run the ML algorithm. In principle, knowledge of medical diagnosis can be derived automatically from the information of the past solved cases[19]. Then the derived ML model can be used to help the physician when facing new patients to improve the speed, accuracy, and reliability of the diagnosis[19].

There are many successful applications for the medical diagnosis of various diseases. The most widespread application of AI in medical diagnosis is in the field of radiology, due to the significant advancement of image recognition tasks[20]–[22]. Many research teams have developed image processing and computer vision algorithms to achieve better diagnosis[23]–[25], enhanced visualization of pathology[26]–[30], alert on emergencies[26]–[28]. Zhang et. al. designed an AI system that can diagnose

COVID-19 pneumonia according to the CT scans and can improve the performance of junior radiologists to the senior level[31]. Becker et.al. and Chougrad et.al. conducted DL methods to screen for breast cancer to help the radiologist classify mammography mass lesions, and they have proven to be accurate in identifying signs indicative of breast cancer and predicting whether the lesions are benign or malignant[32], [33]. Ogino et. al. conducted DL-based medical image analysis on the pathological tumor classification of prostate cancer and found that it achieved a high classification accuracy given only radiological images as input, which can significantly help improve the diagnostic prediction performance of radiological images[34]. Besides the radiology, AI also has been applied in the field of oncology[35]–[37], cardiology[38]–[40], gastroenterology[41], [42], and ophthalmology[43]–[45]. Firstly, for oncology, McKinney proposed an AI system that can surpass human experts in breast cancer prediction and found that the AI system maintained a good performance across different datasets from the United Kingdom and the United States[46]. This robust evaluation of the AI system paves the way for clinical trials to improve the efficiency of breast cancer screens [46]. Wang et. al. proposed an AI tool that is consistent and even often better than most of the experienced expert pathologists in diagnosing colorectal cancer using weakly labeled pathological whole-slide image patches[47]. As for cardiology, Attia et. al. developed an AI-enabled ECG using CNN to detect atrial fibrillation during normal sinus rhythm using standard 10-second, 12-lead ECGs[48]. Wu et. al. employed an ANN approach to predict myocardial infarction efficiently, which can provide valuable

insights for reducing misdiagnosis in clinical trials[49]. Tan et. al. applied heart sound signal processing and CNN for analysis and classification of congenital heart disease heart sounds, which can effectively improve the robustness and accuracy of heart sound classification and can be further applied to machine-assisted auscultation[50]. AI also has successful application in the field of gastroenterology, such as detecting inflammatory bowel disease[51], [52], ulcerative colitis[53], gastrointestinal bleeding[54], [55], atrophic corpus gastritis[56], and gastroesophageal reflux disease[57]. Then, for the diagnosis of eye diseases, a lot of studies of AI application in diagnosing ophthalmological diseases have been reported. Most of the works are on diabetic retinopathy[58], [59], glaucoma[60], [61], age-related macular degeneration[62]–[64] and cataract[65]–[67], which are the four top leading cause of adult blindness. In addition to disease diagnosis, AI is also applied in surgery[68], [69]. Lots of studies have proved that AI can be used to process large amounts of surgical data to identify or predict adverse events in real-time and further support intraoperative clinical decision-making[70]–[73].

AI also has successful applications in the field of mental health to detect mental disorders, such as depression. A lot of researchers used data from social media to identify users who are at risk of depression[74]–[76]. The AVEC, which aims to detect depression using affective computing, has been successfully held several times[77]–[80]. Using machine learning to diagnose depression is getting more and more attention. Many participants actively shared and published their latest research results in these

competitions[81]–[83], greatly promoting the development of the automated depression detection system. In addition, many researchers collected their dataset and tried to find more representative depressive symptoms and built automated depression detection systems based on that[84]–[86].

AI also can be used for TB diagnosis. Some researchers used TB chest radiographs with the help of CNN to diagnose TB[85]–[87]. Metabolomics analysis can be excavated to find biomarkers that are useful for initial screening and diagnosing, which can be an alternative method for TB diagnosis. Metabolomics analysis can be divided into targeted metabolomics and non-targeted metabolomics. Targeted metabolism is the detection of specific metabolites, which can achieve absolute quantification of target metabolites. Non-targeted metabolomics detects all detectable metabolite molecules in a sample unbiasedly. Many metabolomics studies have found sputum[88], [89], blood[90], [91], breath[92], [93], and urine[94] can be used for identifying new biomarkers for TB infection or treatment response.

In medicine, AI can improve patient healthcare through earlier detection and diagnosis, and improve workflows, thereby reducing medical errors, lowering medical costs, and finally lowering morbidity and mortality.

1.3 Gene Therapy and Gene Delivery

After the disease diagnosis or prognosis, the next step is disease treatment. Gene therapy is a technology that transfers genetic material into specific cells of a patient to

treat or cure the disease[87]. It can work by several mechanisms: replacing the disease-causing gene with a healthy copy of the gene; inactivating dysfunctional disease-causing genes; introducing a new or modified gene into the patient's body to help treat diseases[88]. Gene therapy has been applied in several kinds of genetic diseases, including hemophilia[89], muscular dystrophy[90], and cystic fibrosis[91]. By transferring genes to increase naturally occurring proteins, to change the expression of existing genes, gene therapy can also be used in cardiovascular diseases[92], neurological diseases[93]–[95], infectious diseases[96], and cancers[97]–[99]. Gene therapy requires the identification of therapeutic genes and efficient transfer of the genes to targeted cells. Although short-term gene expression is sufficient for some applications (such as cancer treatment), long-term gene expression is required in most cases[100]. Moreover, it is essential to strictly regulate gene expression levels. Finally, the toxicity and pathogenicity of the delivery vector and the immune response must be considered[100]. The main limitation of the development of human gene therapy is still a lack of safe, effective, and controllable methods for gene delivery[101].

Gene delivery systems are essentially necessary for the gene therapy of genetic diseases[102]. It is now a popular research field with significant demand and can be used in both clinical and scientific biomedical research[103], [104]. There are many successful applications for helping gene therapy. As we know, the selectively permeable plasma membrane can protect the mammalian cells from external environments. Therefore, it is important to find an effective method to transfect cells for gene delivery.

In terms of the mechanism of delivering genetic material through the cell membrane o the nucleus, two methods can achieve the gene delivery: increasing the cell membrane permeability and thus promoting the penetration of the target gene, or developing an ideal carrier which gets low cost, high loading capacity, stability, no or low toxicity and easy operating to carry the target gene to pass through the cell membrane and release it to the nucleus[105]. There are many available gene delivery methods, viral-vector system approach is the most widely used method now[108], which can achieve high transfection efficiency with the safety issues related to immunogenicity as its main weakness[107]. So, there is a lot of work on the non-viral approaches, including liposome-based methods[109], calcium phosphate precipitation[110], cationic polymers[111]–[113], and nanoparticle-based hybrids[114]. In addition to these chemical approaches, physical delivery methods are attracting more and more attention of researchers, including the electric field[119], the acoustic method[120], and physical injection[121], to disrupt the cell membrane to increase the permeability and enhance the gene delivery. Poor transfection efficiency and human used concerns are always the problems, so the search for a new or combined method that could improve the gene delivery significantly remains a big challenge for researchers.

1.4 Ultrasound

The ultrasound method is one of the widely used acoustic transfection methods mentioned above and has been proved to be an effective method of delivering genes

into cells and tissues[123]–[127]. Ultrasound is an acoustic wave characterized by frequencies greater than 20 kHz, which is beyond the limit of the human hearing range[128]. Ultrasound has been used for military and industrial applications for a long time in the early decades. Besides these applications, ultrasound now is widely used for diagnosis, surgery, and therapy[128], [129]. Sonography is the most common diagnostic application of ultrasound, and it can visualize and monitor the internal tissue of our body by using an external detector pressing on our skin[130]. At its early therapeutic applications, researchers focused on the treatment produced by using the thermal effects of ultrasound, the ultrasound is absorbed by the tissues and converts to heat energy to increase the temperature in target tissues and kill the disease cells. While nowadays more researchers are paying attention to the non-thermal characteristic, including acoustic cavitation and mass transfer enhancement[131]. When ultrasound waves pass through target tissues, they can induce acoustic streaming and cavitation, which can change the concentration gradient, thus enhancing cell permeability. For ultrasound medical applications, the intensity is one of the most important parameters, and the safe intensity range is between 0.05 W/cm^2 and 100 W/cm^2 [132], [133]. Compared to low-frequency ultrasound, higher frequency (1-3 MHz) has been used for widely used as therapeutic ultrasound and has several advantages to penetrate through membranes and minimize the damage to cells at the same time[127]. LIPUS, as a particular type of ultrasounds, is widely used in many aspects of medical applications, such as bone healing[134], inflammation inhibiting[135], soft-tissue regeneration[136],

and the induction of cell-membrane porosity. It generates at a frequency of 1-3 MHz and repeats at 1 kHz with a duty cycle of 20% to outcome a low-intensity pulsed-wave [137]. LIPUS has also been proven to help many types of cells divide and proliferate under the safe operational intensity range of LIPUS (between 0.02 and 1 $W/cm^{2)}$ and treatment durations of 5 - 20 minutes per day, such as algal cell[138], [139], stem/progenitor cell[140] and mammalian cell[141] and can also promote CHO cells growing and antibody production[142], increase cell permeability[138] and enhance gene delivery through the use of microbubble[143]. LIPUS has almost no thermal effects due to its low intensity[144]. Therefore, there is great potential for using LIPUS in gene delivery.

1.5 Contribution and Novelty of This Thesis

Diseases can affect people not only physically, but also mentally. Therefore, disease diagnosis, prognosis, and treatment are crucial for improving human life quality. In this thesis, we proposed biomedical engineering applications that can contribute to both disease diagnosis and treatment. Firstly, two different automated diseases diagnosis approaches are presented and implemented. These two automated diseases diagnosis approaches are based on different ML algorithms and both lead to satisfying results. For the automated depression detection and assessment, 1) Compared to previous studies, few of them employed all of these modalities, text, audio, and video. I propose multimodality features that can capture depression behavior cues from all

these three modalities in a fusion framework. 2) Currently, most of the work for text analysis is based on word embedding. However, this approach gives out poor performance on the long interview sequence. In this case, sentence embedding is used in the depression detection task for the text modality and leads to a great improvement compared to the common word embedding methods. 3) Instead of feeding the low-level descriptors directly to the ML algorithms, we first conducted dimensionality reduction approaches and then used the features along with ML algorithms to predict depression from audio and video, leading to better results. 4) By fusing unimodality predictions, the final fusion model outperformed the baseline no matter on the binary classification task or the regression task. The baseline model was provided in AVEC 2016 using the linear SVM to classify depression and estimate the depression severity from audio and video modalities[78]. For the automated pulmonary TB diagnosis, 1) Compared to the previous reports that just used IDO as a novel TB biomarker or conducted non-target metabolism study, I proposed Kyn, Trp, IDO, and T-spot can all contribute to the TB detection and lead to satisfying performance. 2) Most of the previous work just got their results on the test set and didn't collect independent cohort data to verify their results. In this thesis, independent cohort data is collected to conduct the double-blind experiments to further verify the performance. 3) Currently, the automated TB detection methods are just focused on diagnosing TB from healthy people, few of them also included LTBI and none of them included NTB patients. However, LTBI and NTB patients also play an important role in the clinical TB detection application. In this thesis,

we also collected LTBI and NTB samples and conducted experiments to diagnose ATB from them.

For the gene delivery system, 1) LIPUS gets wide applications in the medical field, however, no one has used it on enhancing gene delivery. I proposed that LIPUS can promote cell growth and enhance cell membrane permeability, therefore improving gene delivery. 2) MNPs, guided by a magnetic field, can further improve targeted gene delivery with minimal side effects on other tissues compared to other conventional approaches; 3) Instead of just using chemical or physical gene delivery methods like most studies conducted, I synergistically combine LIPUS and magnetic fields for gene delivery, which can leverage each method's advantages, therefore achieve a high transfection efficiency with low cytotoxicity and outperform the standard transfection technique on the market.

1.6 Thesis Outline

This thesis is organized into five chapters to present the two automated diseases diagnosis approaches and a gene delivery system. Chapter 1 explains the reason why automated diseases diagnosis approaches and gene delivery systems are indispensable and reviews the research progress from the state-of-the-art. It also presents the contributions and novelty of this thesis. Chapter 2 focuses on depression detection and assessment. It first introduces the current research on automated depression detection and then describes the proposed multimodal fusion model comprised of text, audio, and

video for both depression detection and assessment tasks. Chapter 3 is the automated TB diagnosis using logistic regression models based on the metabolomics results from high-resolution LC-MS. Chapter 4 covers the gene delivery system, that combined the LIPUS stimulation and MNPs, with the help of an external magnetic field. Finally, Chapter 5 summarizes the above research work and presents possible future work.

2 A Multimodal Approach for Detection and Assessment of Depression Using Text, Audio and Video

2.1 Introduction

Depression, also known as major depressive disorder, is presently one of the most common mental disorders and can last for a lifetime. Today, more people than ever are suffering from depression, which may be caused by psychological and social stress[145]. Depression is characterized by impairing patients' ability to cope with stressful life events and usually leads to constant sadness and loss of interest[145]. It can affect people's feelings, thoughts, and behaviors and cause various emotional and physical problems, such as weight loss, insomnia, fatigue, and, in the worst-case, lead to suicide[146]. According to the WHO report on depression in 2017, over 300 million people are estimated to suffer from depression globally[147]. There are psychological and pharmacological treatments for moderating and controlling depression. However, in low- and middle-income countries, treatment and support services for mental health are underdeveloped and an estimated 76-85% of people cannot receive the necessary treatment in these countries[145]. One study shows that the lifetime risk of depression is over 12% for men and 20% for women in the United States[148]. More than 0.7 million people die by suicide each year and the total number of suicide attempts is estimated to be about 10 million per year[149]. Patients tend to self-harm or even

commit suicide in severe cases of depression, and it is estimated that up to 50% of suicides have been clinically diagnosed with depression[150]–[152]. However, statistics show that less than 70% of all patients afflicted by conservative depression in the world would take the initiative to consult a medical practitioner and receive effective treatment to relieve depression[153]. Due to the suffering inflicted on patients and the untimely diagnosis and treatment for depression, the screening test of depression in the early stage is extremely important.

The current mainstream diagnostic methods mainly rely on scales, questionnaires, etc. PHQ-8, HAM-D, and BDI are commonly used questionnaires in the clinical diagnosis of depression. This diagnosis is based on a practitioner's judgment of patients' responses to the questionnaire. Human bias in subjective judgment may lead to misdiagnosis[154]. These methods have limitations on strong subjectivity and low flexibility, resulting in a low accuracy in the diagnosis of depression. Therefore, there is a strong need for an accurate and simple method to detect depression. Recently, with the rapid advancement of artificial intelligence, researchers have found that behavioral cues such as semantics, prosodic features from speech, and facial expressions are feasible candidates for detecting depression[75], [85], [155].

Although researchers hope that AI can contribute to the diagnosis and treatment of depression, the traditional centralized ML requires the aggregation of patient data. However, the data privacy of mentally ill patients requires strict confidentiality, which hinders the clinical application of machine learning algorithms. For data privacy, all the

participants have signed the informed consent and know that the data will be used for research work. Besides, current data collections remove the personal information and just reserve the data correlated to depression, which greatly protects the privacy of patients.

In this thesis, we present a novel and effective multimodal framework that extracts important features to achieve two tasks on different modalities. The first task is to diagnose depression (binary classification task) and the second is to predict the degree of depression (regression task), which is measured by PHQ-8. The primary contribution and novelty lie in proposing multimodal features that can capture depression behavior cues in a fusion framework. A sentence embedding that was originally proposed for calculating the similarity of pairs of sentences is used in the depression detection task for the text modality and leads to a great improvement compared to the common word embedding methods. Low-level descriptors and SVM along with PCA are proposed to predict depression from audio. On the other hand, we investigate the impact of different feature sets on video modality and show that the XGBoost gives us the importance of each feature and conducts the feature selection based on the cross-validation, leading to better results. We implement a late decision fusion layer to integrate different modalities which help identify depressed subjects with appreciable accuracy as evident in results obtained. Meanwhile, we introduce an over-sampling strategy to train our model, which greatly alleviates the bias caused by the imbalance in the data distribution. The network shown in Fig 2.1 utilizes several features from the text, audio, and video

modalities and employs various machine learning models. We performed several experiments to obtain and combine the best machine learning models for each modality. Our final fusion model achieved a high weighted F1 score and low RMSE, MAE, which outperformed the baseline on both the binary classification task and the regression task[78].

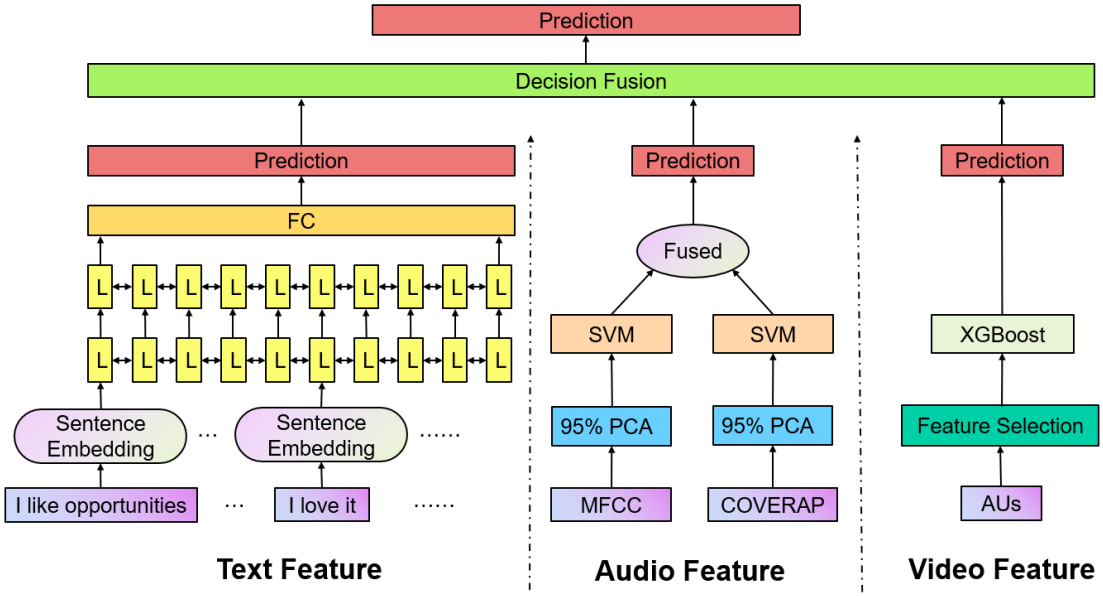


Figure 2. 1 Block diagram of proposed network on multi-modality input features.

2.2 Related Work

This section presents an overview of the current state-of-the-art related to the automated depression detection method proposed in this thesis.

2.2.1 Text

Text, more specifically the verbal content of the words a person speaks is an important feature for detecting depression[156], [157]. With the surge in the use of

social media, a large amount of text data is produced and shared on social media, which provides researchers the opportunity to analyze depression from the text. This data can help with sentiment analysis and provide insights into the relationship between their social media content and mental states[76], [158] due to the anonymous and open nature of social media, individuals share real emotions that represent their real mental states[74]. Ansari et al. proposed a Markovian model to detect depression using content rating provided by human subjects[159]. Tong et al. employed a novel classifier inverse boosting pruning tree to detect depression from online social behaviors[160]. Orabi et al. implemented word-embedding and deep learning methods to detect depression for Twitter users[161]. Islam et al. used decision tree techniques and identified high-quality solutions to mental health problems among Facebook users[162]. However, the social media data has some common limitations, the noise behind the original post, such as acronyms, buzzwords, etc., makes data preprocessing very difficult. In addition, people are more likely to generate negative content on social media because they are anonymous, although they are healthy. Besides the social network data, there are also lots of articles on existed depression corpus. Yin et al. applied semantic embedding and emotional embedding along with the hierarchical RNN to predict depression[163]. Niu et al. matched questions and answers to QA-pair and further employed a pre-trained GloVe word embedding and graph attention model for depression detection[164]. Hanai et al. used Doc2Vec to generate word embeddings and fed them into a BiLSTM neural network to diagnose depression[165]. For most social media data, text analysis was

based on short texts and these classifiers do not perform well in long conversation text data.

2.2.2 Audio

Although the relationship between verbal content and depression severity level is more prominent, speech audio features, known as prosodic and acoustic features, also play a pivotal role in diagnosing depression[166]. Clinicians treat audio features, such as reduced rhythm, less or monotonous speech activity, and energy in speech as important signs of diagnosing depression, and audio is a kind of easy recording feature. These two main reasons make speech audio critically popular in the automatic depression detection topic[166]. Commonly used audio features are spectrograms, power, MFCCs, and deep spectrum representations[167]. Recently, deep neural networks have been used to extract discriminative features from speech. Due to their data generalization capabilities, they can learn more robust data representations[168], [169]. Some researchers used deep spectrogram images as input and employed CNNs to extract important features to further predict depression[163], [167], [170]–[172]. Mel-spectrogram is a spectrogram converted from frequency to Mel scale, which is more suitable for applications because it is a perception scale of pitches judged by the listener to be equal to each other and has shown strength in depression detection topics. Many approaches employed Mel-spectrogram and CNNs to predict depression[171], [173]–[175]. Pampouchidou et al. used statistical descriptors of several pre-extracted

audio features and the decision tree to assess depression[176]. Nasir et al. generated a new vector (i-vector) representing the speaker acoustic model on MFCCs for depression detection based on the Gaussian mixture model-universal background model[146]. Similar to the MFCCs sequences, reports have shown that the raw features sequences extracted from COVAREP[78] are potentially effective at predicting depression[154], [163]–[165]. However, the spectrogram-based deep learning methods have their limitation. The spectrogram with x-axis as time and y-axis as frequency is a position-sensitive plot, while CNNs do not take into account the position information from the spectrogram. As for long sequence deep learning methods, the sequence can be extremely long, so the deep learning models may face challenges like slow inference, vanishing gradients, and difficulty capturing long-term dependencies. Due to the limited training sample size, applying deep learning methods tends to lead to overfitting, hence we avoided deep learning methods on audio features.

2.2.3 Video

In addition to text and audio, video features play a key role in modeling the deep correlation between depression and facial emotions. It has been observed that patients with depression often display distorted facial expressions, such as twitching eyebrows, sluggish smiles, frowning faces, aggressive expressions, restricted lip movement, and reduced blinking times[154]. MHI is an algorithm widely used in the field of human action recognition[177]. Meng et al. proposed using MHIs along with extracting LBP

and Edge of Oriented Histograms for depression recognition in the AVEC 2013 Challenge[77], [178]. Niu et al. utilized the raw video segments with 3D CNN and LSTM to predict the BDI score of patients[82]. The authors presented OpenFace an open-source interactive tool to estimate facial behavior[179]. OpenFace is a widely-used tool, which provides features for face landmark regions, head pose estimation, eye gaze estimation, and facial action unit. Many approaches related to facial expression are based on this tool. Pampouchidou et al. also first extracted 68 two-dimensional landmarks from the raw video and selected facial landmarks affected by smiling to feed into the nearest-neighbor model[180]. Wang et al. fed 68 two-dimensional landmarks to the LSTM network for depression diagnosing[181]. The main concern for these methods is that using features of every frame would be very tedious and would be limited by the information we can extract from the long interviews. In this thesis, we used the low-level video features extracted from OpenFace.

2.2.4 Multimodality

Multi-modal methods integrated text, audio, or video features to detect depression. Kächele et al. integrated visual and audio features to explore the depression state through a hierarchical classifier system[182]. Chao et al. extracted visual and audio features and used them to train an LSTM to learn long-term temporal dynamic features and applied multi-task learning to optimize the prediction of depression severity[183]. Ringeval et al. designed a hybrid depression estimation framework using audio, video,

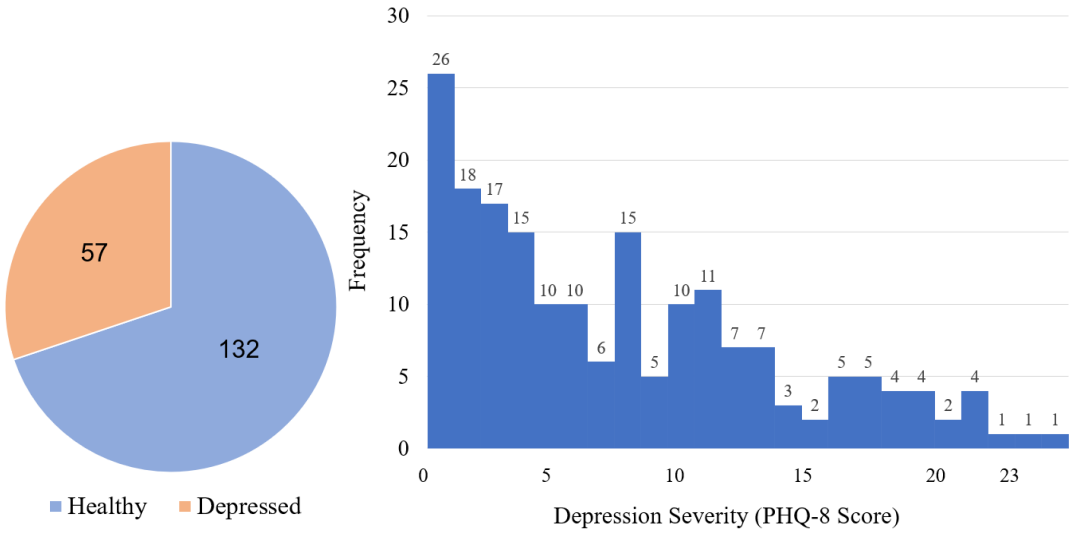
and text descriptors to predict depression[79]. Gong et al. utilized audio, video, and semantic features to propose a novel topic-based modeling method for context-aware analysis of depression[83]. Most of them just utilized one or two modalities. In this thesis, we also performed a multimodality model using text, audio, and video features and outperformed our unimodal models.

2.3 Data

2.3.1 Dataset

In this thesis, we adopted the DAIC-WOZ dataset[184] for training and testing. This dataset is publicly available and also used for the depression sub-challenge task in the AVEC 2016, 2017, and 2019, requiring participants to use audio, video, and text analysis to predict depression[78]–[80]. All the user have singed the agreement form to protect the patient privacy. All the The dataset includes data from 189 subjects. For each subject, the dataset includes the raw audio and transcript of an interview ranging between 7-33 minutes. During the interview, an animated virtual interviewer, called Ellie, controlled by a human interviewer in another room takes the initiative to ask questions designed to support the diagnosis of depression. Besides the raw audio of the interviews, the dataset also contained some baseline features both for audio and video. The dataset comprises two target labels: PHQ-8 binary labels (PHQ-8 scores ≥ 10) and PHQ-8 scores for each participant. The data distribution is shown in Fig 2.2. There is a

total of 189 subjects in the DAIC-WOZ dataset, out of which 107 subjects are used for training, 47 for testing, and 35 for development. Since the test labels are not available in the challenge, most of the literature reported their performance on the development set. To better compare the performance with others, the results shown in this thesis are evaluated on the development set.



(a) (b)

Figure 2. 2 Depression and severity level distributions of the participants within the DAIC-WOZ corpus. (a) The number of individuals in Depressed and Healthy groups. (b) The histogram of Depression Severity across the twenty-four depression severity levels given by the PHQ-8 test.

2.3.2 Data Augmentation

An imbalanced dataset is a dataset with a severely skewed class distribution. This imbalance could affect the performance of many machine learning algorithms, leading

to ignoring the minority class entirely which may result in overfitting. This is always a big issue because it is typically the predictions on the minority class that are most important. As we see in Fig 2.2, the DAIC-WOZ dataset presents an unbalanced number of samples in different groups. The number of healthy participants is almost three times higher than that of depressed patients. To increase the number of depressed subjects and improve the performance and robustness of the models, we utilized the re-sampling method on the dataset to obtain balanced data. In our case, we over-sampled the subjects from the minority class (depressed class) to make a balanced distribution of classes on the training set. For the development set, we kept its original distribution.

2.4 Methodology

This section presents the models implemented for depression detection. A block diagram of the overall proposed multimodality with a late decision fusion network is shown in Fig 2.1. Sections 2.4.1 to 2.4.3 explore each single modality representation for feature extraction and the depression detection task. Section 2.4.4 presents a model for the decision fusion layer by combining the single modalities for both classifying depression and predicting severity.

2.4.1 Text Model

For the text modality, we used the speech-to-text transcripts provided in the dataset[184]. An example of a participant's transcript data is shown in Table 2.1.

Because we aim to detect depression for participants and the questions from Ellie are similar among different interviews, we removed Ellie's parts and processed the participants' responses in succession. The raw text was initially preprocessed. Since some participants used colloquial English words, we manually modified the utterance by replacing these words with the original complete words. Otherwise, these words become all out of vocabulary words while training the neural network during the training and affect the model performance. Meta information such as <laughter> or <sigh> can be helpful for model training, so it was retained by removing the angle brackets. Moreover, we removed the stop-words and tokenized the remaining transcripts.

Context-free word embeddings are commonly used in text classification tasks, either trained from scratch with Word2Vec[185] or a simple pre-trained word embedding GloVe[186]. However, these word embedding methods aim to capture the context of a specific sentence by only considering surrounding words. Therefore, it does not capture the meaning of the sentence. Modern word embeddings like BERT (Bidirectional Encoder Representations from Transformers), built with self-attention mechanisms and LSTMs are context-sensitivity, which means they will produce sentence-level representations. Since depression data is difficult to obtain, using a pre-trained model on a large text body can help alleviate the problem of data sparseness. The BERT model is pre-trained on a large corpus, making them effective advanced feature extractors[187]. The BERT model generates embeddings for words based on the

context in which they appear, thereby producing slight changes for each word appearance. This requires the entire sentence to generate word embeddings. Therefore, they are fundamentally different from traditional Word2Vec embeddings, which create sentence embeddings as the average of all word embeddings. In our case, the maximum length of our transcript is 2543 words. Each word contains little information about the entire context. Using word embedding methods can result in lots of important information loss during training. Instead of using word embedding, we utilized sentence embedding in extracting text features. Sentence embedding can be extracted at multiple levels such as characters, sub-words, words, sentences, and paragraphs. It can represent the entire sentence and its semantic information as a vector, which helps understand the context, intention, and other nuances in the entire text. At the same time, the input time step is greatly reduced, which is very helpful for training. Traditional word embedding represents each word by a D dimensional vector. For the sentence embedding, the general representation for a sentence j is extracted from the average of their word-level representations w_t as:

$$x_j = \frac{1}{N_j} \sum_{n=1}^{N_j} w_t$$

where N_j is the number of words in that sentence. In our work, we employed a pre-trained model. Five different embedding methods are tested: GloVe word Embedding[186], FastText word embedding[188], BERT sentence embedding[189], InferSent sentence embedding[190], and Universal sentence encoder (USE)[191].

After embedding words/sentences, we padded the inputs with zeros for a constant input tensor size. All utilized embedding methods with their dimension D are shown in Table 2.2. We used two layers of stacked BiLSTM network architecture with the input of embeddings. LSTM is a type of recurrent neural network capable of modeling sequence-dependent data in sequence prediction problems. The BiLSTM can learn both forward and backward directions effectively increasing the amount of information available to the network, improving the context available to the algorithm. Each BiLSTM layer has 512 hidden units, where the output of each hidden unit of the first BiLSTM layer is the input of each hidden unit in the second layer. Then we concatenated the last states of both forward and backward directions and fed them to a fully connected layer with one output node. The overall text modality network is shown in Fig 2.1. For the binary classification task to diagnose depression, we employed an activation layer with sigmoid function and used BCE loss function for training:

$$BCELoss = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\big(p(y_i)\big) + (1 - y_i) \cdot \log\,(1 - p(y_i))$$

where y is the label (1 for depressed and 0 for healthy in our case) and $p(y)$ is the predicted probability of the subject suffering from depression.

While for the regression task to predict PHQ-8 scores, we removed the sigmoid activation layer and applied the MSE loss function for model training:

$$MSELoss = \frac{1}{N}\sum_{i=1}^{N} (y_i^{true} - y_i^{pred})^2$$

where y^{true} is the actual PHQ-8 score and y^{pred} is the predicted PHQ-8 score for this subject.

Table 2. 1 An Example of a Random Participant's Interview Transcript.

Start Time	Stop Time	Speaker	Value
49.672	50.132	Ellie	How are you doing today?
53.440	57.160	Participant	Today is a wonderful day I am doing just absolutely marvelous.
57.892	59.252	Ellie	That's so good to hear.
60.088	61.408	Ellie	Where are you from originally?
61.490	67.980	Participant	I was born in Cleveland; I was raised in Tucson Arizona and came to Los Angeles when I was sixteen years old.

Table 2. 2 Embedding Dimensions Utilized.

Level	Embedding	Dimension
Word	GloVe	100
	FastText	100
Sentence	BERT	1024
	InferSent	4096
	USE	512

2.4.2 Audio Model

For the audio modality, we created models using different audio features. The dataset provided pre-extracted features using the COVAREP toolbox[192]. Audio features are sampled at 100 Hz (every 10 ms) over the entire recording. The features contain prosodic features, voice quality features, and spectral features, including F0, VUV, NAQ, QOQ, H1, H2, PSP, MDQ, peak/slope, Rd, Rd conf, MCEP 0-24, HMPDM 0-24, and HMPDD 0-12. VUV is an important feature that provides a label for the

current segment as voiced or unvoiced. In the unvoiced segment (VUV=0), there is no vocal sound detected, so all the features measured for this segment (i.e., F0, NAQ, QOQ, and H1H2) are meaningless and cannot be further utilized. First, similar to the text feature, we removed all the frames of Ellie and isolated the participants' voices according to the timestamps in the transcripts. Next, frames with unvoiced segments (VUV=0) were removed from the final concatenated time series. The MFCC is the cepstrum parameter extracted in the frequency domain of the Mel scale, which describes the non-linear characteristics of human ear frequency. It is widely used in speech recognition due to its robustness and is comparable to the auditory characteristics of the human ear. In addition to COVAREP features, we extracted 13-dimensional MFCCs every 10ms, and first (Δ) and second-order differentials (ΔΔ) of the MFCCs as features. Using the two sets of audio features: COVAREP and MFCCs, we calculated the statistical features for these low-level descriptors shown in Table 2.3. We did not utilize the raw sequence of these audio features because the average and maximum of the participants-only frame length are 24556 and 69182, respectively. Similar to the text feature, such long frames can lead to a significant increase in training time. In addition, trying to perform backpropagation in a long sequence may cause the gradient to disappear, which in turn will weaken the reliability of the model. Moreover, because the overall sample size is only 189, using a deep learning model easily leads to over-fitting. Therefore, we extracted representative statistical features and fed them into machine learning models.

Table 2. 3 Statistical Descriptors Calculated from Two Sets of Audio Features.

Low-level Descriptors	Statistical Descriptors	Dimensions
F0, NAQ, QOQ, H1H2, PSP, MDQ, peak slope, Rd, Rd conf, creak, MCEP 0-24, HMPDM 0-24, HMPDD 0-12	mean, max, min, skewness, kurtosis, standard deviation, median, root mean square level, peak-magnitude to root-mean-square ratio, interquartile range.	73*10
MFCC, Δ, ΔΔ		39*10

Because the dimensions of the audio features, 730 and 390, are both larger than the sample size, 189, we need to reduce the input dimension, otherwise, it will result in a dimensional disaster. There are two ways for dimensionality reduction, feature extraction, and feature selection. Feature extraction aims to reduce the number of features in the dataset by creating new features from existing features, while feature selection is to select input variables that have the strongest relationship with the target. In this work, we first applied PCA to reduce the dimension of the input features. PCA is commonly used to reduce the dimensionality of a large data set, by converting a large group of variables into a smaller set of variables and can minimize the loss of original information. We used 95% of the variance for PCA projection in the audio modality for dimension reduction. After this, we achieved 84 and 72 dimensions for COVAREP and MFCCs feature sets, respectively. We then fed these features into an SVM to get an early prediction. SVM is a supervised machine learning algorithm that can be used for classification or regression problems, which uses a technique called kernel tricks to

transform data and find the best boundary between possible outputs based on these transformations. We then used an AND/Average gate to fuse the early prediction from two feature sets to get the final prediction of audio modality for the classification and regression tasks.

2.4.3 Video Model

Since no raw video is publicly available in the DAIC-WOZ dataset, we can only utilize the video features provided by the dataset. It contains the 2D and 3D coordinates of the 68 facial landmarks, 20 FAUs, the gaze direction and position of eyes, and the position and orientation of the head, for each video frame, along with the timestamp, confidence weight, and detection-success. These features are computed from the raw video using OpenFace[179]. The 68 facial landmarks are points on a human face shown in Fig 2.3 that localize the regions around the eyes, eyebrows, nose, mouth, chin, and jaw. Since the 3D coordinates of the 68 facial landmarks contain all the information contained in 2D coordinates, we just discarded the 2D coordinate feature. The FAUs are related to several emotions, happiness, sadness, surprise, fear, anger, disgust, and contempt. The gaze feature gives out the gaze direction of both eyes and the gaze in head coordinate space. The last feature, head pose, is the head position and rotation coordinates. In addition to these feature sets provided by the dataset, we extracted certain distances between facial landmark pairs as geometric feature sets, which are affected by smiling because reduced smiling is very significant in individuals with

depression[193]. The distances we used are shown in Fig 2.3 as red lines. To eliminate the difference between different faces, all distances were normalized based on the cheekbone width (landmarks 1 and 17) represented by the green line shown in Fig 2.3. We removed all the frames with detection_success of 0, meaning that these frames have not been successfully extracted facial features. We then calculated the changing speed (Δ) and acceleration (ΔΔ) of the 3D landmarks' positions and the geometric features. Similar to the audio features, we extracted some representative statistical descriptors of these feature sets and fed them into machine learning models. An overview of the video feature set is shown in Table 2.4.

Table 2. 4 Statistical Descriptors Calculated from Video Feature Sets.

Features	Statistical Descriptors	Dimension	Dimension After Feature Selection
FAUs	mean, max, min, skewness, kurtosis, standard deviation, median, root mean square level, peak-magnitude to root-mean-square ratio, interquartile range.	20*10	75
3D Landmarks		204*10	48
3D Landmarks Δ		204*10	123
3D Landmarks ΔΔ		204*10	28
Head Pose		6*10	10
Eye Gaze		12*10	25
Geometric Distance		10*10	27
Geometric Distance Δ		10*10	19
Geometric Distance ΔΔ		10*10	49

Due to the high dimensionality of the video feature set and the attendant risk of overfitting, we performed feature selection before further prediction to select a reduced

feature subset. We used XGBoost to achieve feature selection and the depression prediction together. XGBoost is an ensemble machine learning algorithm based on decision trees, utilizing a gradient boosting framework. One advantage of using gradient boosting is that after building the boosted tree, it retrieves the importance score for each feature. Generally, importance provides a score that indicates the usefulness or weight of each feature in building a boosted decision tree within the model. The more attributes that use decision trees to make key decisions, the higher their relative importance. We first applied 5-fold cross-validation on the training set to select an optimal importance threshold and used the attributes with higher importance as a new feature set. Then we employed the same feature subset on the development set and further fed it into the XGBoost to evaluate the performance on diagnosing depression state and severity.

2.4.4 Fusion Model

We first obtained classification/regression depression prediction for text, audio, and video modalities respectively. After we got the prediction for each modality, we performed late-fusion of decisions from multiple modalities to investigate its effectiveness at predicting depression. We utilized a voting ensemble to fuse the predicted results from various modalities for both classification and regression tasks. In classification, the voting ensemble involves summing the votes for crisp class labels from all modalities and predicting the final class with the most votes. For regression, a

voting ensemble is involved in predicting by taking the average of the predicted PHQ-8 scores of the multiple modalities' models.

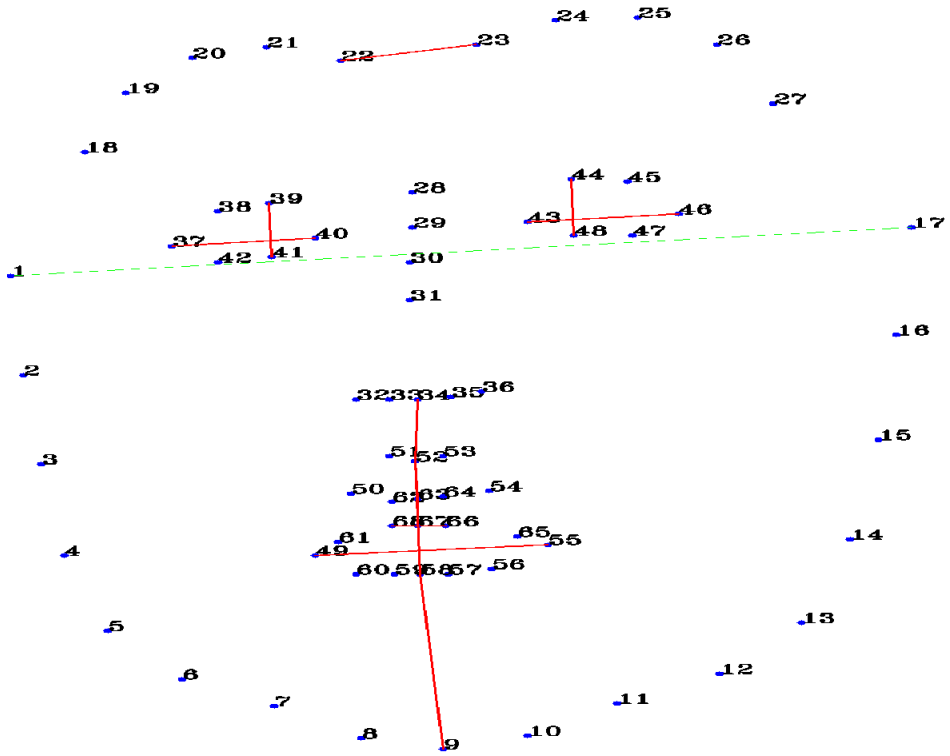


Figure 2. 3 68 2D Facial Landmarks and 10 Geometrical Features.

2.4.5 Evaluation Metric

There are two tasks in this work, classification for diagnosing depression and regression for predicting the degree of depression. For this, we applied two different evaluation metrics. For classification, we utilized the F1-score for healthy and depressed classes to measure the performance. Since the dataset has a skewed distribution of classes, we also provided the weighted F1 score. The weighted F1 score is calculated by averaging the class-specific F1 scores scaled by the relative number of

samples from that class. The regression performance was evaluated using MAE, $\frac{1}{n}\sum_{i=1}^{n}|y_i^{true} - y_i^{pred}|$ and RMSE, $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i^{true} - y_i^{pred})^2}$. For both tasks, we reported the performance of single modality and multi-modal fusion methods.

2.5 Result and Discussion

We evaluated the proposed multimodality framework on the development set of DAIC-WOZ. F1 scores for healthy and depressed classes and the weighted F1 scores were calculated for unimodal and multimodal models on the development set. We employed "T", "A", and "V" to represent text, audio, and video modality, respectively. Table 2.5 illustrates the unimodal classification results and the reference state-of-the-art models for each modality in bold font. The same random seed was set for all the experiments to exclude the random influence as the result of dataset split and different initialization. We applied these results to select our unimodal features and further used these features to fuse predictions.

Table 2. 5 Comparison of F1 score for the Single Modality's Classification.

Modality	Feature	F1 score (Healthy)	F1 score (Depressed)	F1 score (Weighted)
T	GloVe	0.680	0.200	0.515
	FastText	0.793	0.000	0.521
	BERT	**0.870**	**0.750**	**0.829**
	InferSent	0.793	0.583	0.714
	USE	0.723	0.435	0.624
A	MFCCs	0.840	0.600	0.758
	COVAREP	0.755	0.235	0.577

	MFCCs+COVAREP	**0.826**	**0.667**	**0.771**
	MFCCs+COVAREP (raw)	0.783	0.583	0.714
V	**FAUs**	**0.833**	**0.636**	**0.766**
	FAUs (raw)	0.792	0.545	0.707
	3D Landmarks	0.760	0.400	0.637
	3D Landmarks Δ	0.750	0.455	0.649
	3D Landmarks ΔΔ	0.784	0.421	0.660
	Head Pose	0.679	0.000	0.446
	Eye Gaze	0.667	0.273	0.532
	Geometric Distance	0.681	0.348	0.567
	Geometric Distance Δ	0.692	0.111	0.493
	Geometric Distance ΔΔ	0.731	0.222	0.556

From Table 2.5, as we expected, the sentence embedding methods (BERT, InferSent, USE) all have better performance than the word embedding methods (GloVe, FastText). This means that compared to word embedding, sentence embedding is more effective and can retain more information for training on long sequence interviews. Our experimental results comparing word-level and sentence-level embeddings are also shown in Table 2.5. For both front-end features, considerable improvements were observed when using sentence-level features instead of word-level features. The text modality model's performance regarding weighted F1 score (0.52 to 0.83) significantly benefited from sentence-level features. We continued to compare sentence embedding methods towards their usage in depression detection. The BERT sentence embedding method significantly outperforms other sentence embedding methods and achieved a weighted F1 score of 0.829 for the text modality. This is because the BERT embedding is pre-trained on large text body datasets and can be effectively used as a high-level

semantic feature extractor. Moreover, it employed an attention mechanism on the LSTM network, which can provide better sentence representations.

For the audio modality, when we just utilized a single feature set, the performance is poor, especially for the COVAREP feature set. We then fused the prediction from MFCCs and COVAREP using a logical AND gate. The model provided improvement to both the healthy and depressed classes, thereby asserting the robustness of the system. The results indicate that the AND gate is effective to fuse predictions from different feature sets for the classification task and can achieve a good result. The best performing audio system achieves F1 scores of 0.826 and 0.667 for healthy and depressed classes, respectively. We also include the results of using raw features without dimensionality reduction. The weighted F1 score decreased 7% compared to the model utilizing PCA to reduce the dimension of the input features. The results illustrate that dimensionality reduction is important for high-dimensional inputs. More specifically, PCA is an effective way for dimensionality reduction in our task and can significantly improve the model performance. The other two common classifiers, XGBoost and KNN were employed with different dimensionality reduction methods. The results shown in Table A.1 illustrate that the proposed framework, SVM with PCA, got the best performance for audio modality.

Table 2. 6 Comparison Between the Proposed Model and other Depression Detection Methods on the DAIC-WOZ Development Set.

Method	Modality	F1 score (Healthy)	F1 score (Depressed)	F1 score (Weighted)	RMSE	MAE
Proposed Model	T	0.87	0.75	0.83	5.61	**4.25**
	A	0.83	0.67	0.77	6.64	5.60
	V	0.83	0.64	0.77	6.45	4.97
	T+A+V	**0.90**	**0.76**	**0.85**	**5.57**	4.48
Baseline[78]	A	0.68	0.46	0.61	6.74	5.36
	V	0.90	0.50	0.76	7.13	5.88
	A+V	0.90	0.50	0.76	6.62	5.52
Alhanai et al.[176]	T	-	-	0.67	6.38	5.18
	A	-	-	0.63	6.5	5.13
	T+A	-	-	0.77	6.37	5.1
Nasir et al.[146]	A	0.89	0.57	0.78	6.73	5.82
	V	0.89	0.63	0.80	7.86	6.48
	A+V	0.89	0.63	0.80	-	-
Rohanian et al.[81]	T	-	-	0.69	6.05	4.98
	T+A	-	-	0.80	5.14	3.66
	T+A+V	-	-	0.81	4.99	3.61

A comparison of the performances of different video feature sets is presented in Table 2.5. As we see, most of the feature sets got poor performance with an F1 score for the depressed class below 0.50. The best performance is obtained with the FAUs features, with F1 scores of 0.833 and 0.636 for healthy and depressed classes, respectively. The FAUs are related to several emotions: happiness, sadness, fear, and disgust, which have a strong correlation with depression. While the rest of the feature sets are not as good as FAUs at providing direct and useful information, it is impossible to capture useful depression-related information only from the statistical descriptors of coordinates and orientations along with the changing speed and acceleration. The results of using raw statistical descriptors of FAUs without dimensionality reduction

are also presented in Table 2.6. Using a model with a selected subset of features with the strongest correlation with depression achieves 8% greater performance compared to a model using raw statistical descriptors. Therefore, XGBoost feature selection can retain important features and eliminate irrelevant features effectively in our task and further improve the model performance. The results of SVM and KNN with different dimensionality reduction methods are shown in Table A.1, showing that the proposed framework, XGBoost with XGBoost feature selection, achieved the best performance for video modality.

From Table 2.5, the proposed best unimodal models shown in bold achieved good performance with an average F1 score of 0.79 on different modalities separately for the classification task. After acquiring the best models for each modality, we utilized the major voting layer to fuse the predictions from each modality by using the selected unimodality models directly for predicting the degree of depression. In Table 2.5, the text unimodality model outperforms the audio and video models, so we assigned more weight to the text model when we performed the weighted average of multiple modalities regression models. The weights for text, audio, and video are 2:1:1 respectively. The results along with other depression detection methods on the DAIC development set are reported in Table 2.6. Dashes indicate the corresponding metric as not reported. From Table 2.6, except for the unimodal text model (not included in the baseline), the audio and video unimodal models outperformed the baseline model provided in the AVEC 2016 challenge[78] when evaluated over the development set.

The proposed multimodal model achieves state-of-the-art performance in both classification metric weighted F1-Score and regression metrics of RMSE and MAE. More specifically, our model achieved a 12% improvement in F1 score compared to the AVEC 2016 challenge baseline[78]. For the regression task to predict the degree of depression, our best models also achieved 16% and 19% improvement in RMSE and MAE, respectively, when compared to the best results in the baseline model. It proves that our model successfully captures the information involved with depression from various modalities automatically. Our model also achieved 10%, 13%, and 12% improvement in weighted F1 score, RMSE, and MAE, respectively, when compared with previous methods[176] which used speech rate, word-level semantic content on text features, and the raw statistical descriptors as input features for audio and video. Our model outperforms 6%, 17%, and 23% in three metrics, weighted F1-Score, RMSE, and MAE when compared to Nasir et al.'s methods[146] which utilized the i-vector on the audio system and random forest on video. Compared to Rohanian et al.'s result using text, audio, and video[81], our regression result is not as good as their best result, because we used the classification task to select features for our two tasks, and our model achieves 5% of improvement on weighted F1 score. The main task in this thesis is to diagnose depression, so depression detection is more important than the PHQ-8 score prediction. The proposed model gives out better performance on the depression detection task. In this case, our model gives out an overall better performance than Rohanian et al.'s results[81]. The above analysis proves that the proposed framework

achieves state-of-the-art performance in both classification metric F1 score or regression metric MAE and RMSE. Our method is also shown to be more robust and effective than others on both classification and regression tasks.

2.6 Conclusion

In summary, I propose a multimodal framework model based on a late decision fusion for depression detection and PHQ-8 score prediction. Firstly, unimodal models that individually consider text, audio, and video features were developed and evaluated independently. The results show that these unimodal models can be effectively used for depression detection separately. For text modality, instead of word embedding, I proposed a novel sentence embedding method to extract semantic representation and greatly improved model performance. For audio and video modalities, we utilized two different methods of dimensionality reduction, PCA and XGBoost, both of which significantly improved the models' performance. These models are then used as highly representative feature extractors and the resulting predictions are combined in a voting ensemble fusion layer. The best results reported in this thesis, weighted F1 score = 0.85, RMSE = 5.57 on the development set, are achieved with a multimodal model that fuses text, audio, and video features. Results indicate that the proposed framework outperforms the baseline models on all five evaluation metrics. The empirical results show that compared with the unimodal model, the employment of the multimodal model provides a better representation, which improves the automatic depression

severity assessment system. This study will contribute to developing automatic depression detection that combines various modalities and can be easily transferred into a high-performance, low-cost, and rapid depression diagnosis and prognosis device.

3 Metabolomic Biomarker Selection for Pulmonary Tuberculosis Diagnosis and Prognosis

3.1 Introduction

TB, derived from the MTB bacterial, is a public health problem in both developing and developed countries and is one of the top 10 major causes of death worldwide[194]. TB results in approximately 10 million new cases of active TB and 1.25 million death in 2018[195]. According to the Global Tuberculosis Report of the WHO in 2019, there are about 1.7 billion people with latent TB infection in the world about a quarter of the population worldwide[196]. The risk of progression from exposure to the tuberculosis bacilli to the development of the active disease is a two-stage process governed by both exogenous and endogenous risk factors[197]–[199]. Socioeconomic and behavioral factors are also shown to increase the susceptibility to infection. Specific groups, such as health care workers and the indigenous population, are also at an increased risk of TB infection and disease[200]. Furthermore, non-adherent behavior with treatment may make this population more contagious[199].

Pulmonary TB occurred when MTB attacked the lung[201]. The lungs are the primary host for MTB infection and TB disease. The traditional symptoms and signs of pulmonary TB include various combinations of cough, sputum, hemoptysis, dyspnea, weight loss, fever, and anorexia[202]. However, these symptoms are not unique to pulmonary TB. Although diagnosis and treatment techniques have been developed for

decades, insufficient case detection and cure rates have been identified as reasons for the increase in the global tuberculosis burden[203]. The diagnosis of TB can now be determined on sputum smear detection, GeneXpert, targeted tuberculin skin test, and the X-ray test[204]. These current tests are not sensitive enough to identify those people who may be infected with TB[205], [206]. Smear detection is the gold standard and gets most-widely used in TB diagnosis but can only be used to diagnose TB when the sputum has enough bacterial load and the accuracy is not stable which can range from 20% to 80% with poor sensitivity and takes four to eight weeks to obtain the results[207]. GeneXpert is more sensitive than smear detection, but it has major limitations of cost and availability, costs far more, requires a continuous power supply, human resources, and expensive operating machine[207]. For the targeted tuberculin skin test, there are also many limitations such as the need for patients to return for the test results and the low specificity because the antigen used for this test is not specific for TB but also present in other cases[208]. X-ray tests require large & expensive equipment. Therefore, finding a low-cost, accurate and rapid method to diagnose TB is a global public health priority.

Metabolism is an unconscious chemical reaction to maintain normal cells and organism function[209]. Metabolites are small molecular intermediate or final products during metabolic processes[210], and thus they are the most abundant representatives of the biochemical activities of organisms. Recently, metabolomics has become a potential tool and has made significant progress in novel biomarker research. It is a

significant challenge to identify sensitive and specific metabolomic biomarkers for accurately diagnosing TB. The analysis of metabolomics methods can be divided into targeted metabolomics and non-targeted metabolomics[211]. The characteristic of targeted metabolomics research is to measure predefined metabolite groups with high precision and accuracy and requires a priori knowledge of the targeted metabolites[212]. Conversely, non-targeted metabolomics research can simultaneously measure a large number of metabolites in each sample, providing more information on relatively quantitative measurement[211]. Chen et al. investigated targeted lipid metabolism, indicating that the selected lipid metabolites could be used as potential biomarkers for TB[204]. Adu-Gyamfi et al. and Suzuki et al. studied Trp and Kyn metabolism to use their ratio IDO to diagnose or predict active tuberculosis disease[213], [214]. Cho et al. performed a targeted metabolomics approach demonstrating that levels of serum metabolites and their ratios could be important indicators for active pulmonary TB[215]. Frediani et al. adopted non-targeted metabolomic analysis, which can differentiate patients with active TB disease[216].

Although many metabolomics studies have found sputum[217], [218], blood[219], [220], breath[221], [222], and urine[223] can be used for identifying new biomarkers for TB infection or treatment response, most studies use multivariate statistical analysis to analyze targeted metabolites and focused on the detection between ATB and HC without an external independent cohort to verify their conclusion. In this work, we first performed targeted metabolomics separately to identify the biomarkers that can be used

to predict patient prognosis. We employed both univariate and multivariate statistical analyses along with machine learning classifiers using Kyn, Trp, and IDO according to the study of Adu-Gyamfi et al. and Suzuki et al. [213], [214] to distinguish between HC, ATB, NTB, and LTBI. To our best knowledge, there are no such studies before trying to distinguish ATB from NTB, and few studies have included LTBI samples. We included the t-spot data and greatly enhanced the ATB vs NTB classification, which is novel and has great clinical medical significance. Finally, we collected an external independent cohort to further verify our results. The results of this study not only confirmed and validated the findings of previous TB metabolomics studies but also proposed novel traditional biomarker candidates for TB based on sufficient data, and got verified on an independent cohort, proved to be further served as a novel method for TB diagnosis and prognosis.

3.2 Material and Methods

3.2.1 Materials

Methanol, acetonitrile, and isopropanol (Optima® LC-MS grade) from Fisher Scientific (Fair Lawn, NJ, USA). Milli-Q® water purification system from Merck Millipore (M.A., USA). Formic acid (M.S. grade) from Fluka, Germany. Ammonium formate (CNW® HPLC grade) and ammonium hydroxide solution (25% NH3, CNW® HPLC grade), and nonadecanoic-d37 acid (C/D/N isotopes®) were obtained from

ANPEL (Shanghai, China). L-2-chlorophenyl alanine was purchased from Intechem Tech. (Shanghai, China). Hexanoyl-L-carnitine-(N-methyl-d3) was acquired from Supelo, Germany. Lysophosphatidylcholine (12:0) was acquired from Avanti Polar Lipids (Birmingham, AL, USA). VACUETTE blood collection tube (Greiner) and frozen pipe (KIRGEN) were used in sample collection and storage.

3.2.2 Ethics Approvals

This study was approved by the Ethical Committee of Shanghai Public Health Clinical Center, and informed consent was obtained from all subjects.

Table 3. 1 Summary of the grouping of samples.

Discovery Set					
Group	**Number of Samples**	**Age**		**Gender**	
		Range	Median	Male	Female
HC	37	19-40	28	33	4
ATB	34	1-85	47	24	10
NTB	35	3-84	60	21	14
LTBI	37	18-18	18	19	18
Total	143	1-85	28	96	47
Validation Set					

Group	Number of Samples	Age		Gender	
		Range	Median	Male	Female
HC	13	18-39	22	11	2
ATB	12	10-71	31.5	10	2
NTB	11	4-80	55	9	2
LTBI	12	18-18	18	7	5
Total	48	4-80	27	38	10

3.2.3 Study Design

In this study, we used high-resolution LC-MS for the metabolomic analysis of plasma samples. I first designed the data collection experiments about the age, gender, and group of the participants to make the data distribution more balanced and diversified. Blood samples of 50 healthy controls, 46 patients with ATB, 46 patients with NTB, and 49 patients with LTBI were acquired. The sex of these groups did not have a statistical difference. The p-value of $\chi 2$ for gender is 0.21. Cause the ages of the patients in our dataset has a large range (1-85), we performed correlation analysis between the concentration of different metabolites and the age of patients. The results shown in Table A.2 indicates that the concentration of different metabolite has a weak correlation with the ages of the patients. We stratified and divided three-fourths of the samples as the discovery set used for training, and the rest of the samples as the

validation set used for verifying. In addition, we further collected 18 samples of ATB and 10 samples of NTB from an independent cohort and used them as external validation to fully prove the effectiveness of our classification model.

3.2.4 Recruitment Criteria

Pulmonary tuberculosis plasma samples were obtained from patients who were pathologically diagnosed with tuberculosis in the Tuberculosis Department of Shanghai Public Health Clinical Center, LTBI were medical staff or family members of confirmed patients, and HC samples were obtained from the Health Examination Center of Shanghai Public Health Clinical Center. All samples were collected between 2018 and 2020. ATB was identified based on sputum or effusion smear or polymerase chain reaction amplification positivity, confirmed by radiological findings and clinical syndromes, alongside a final clinical diagnosis of ATB. LTBI was defined as IGRA positive without clinical syndromes of active TB infection. HC was healthy, no disease at present, no history of tuberculosis, no contact history of tuberculosis, negative laboratory examination and chest X-ray, positive t-spot. The NTB was excluded from the diagnosis of active tuberculosis, including bronchitis, pneumonia, lung cancer, and other respiratory diseases. HIV-negative subjects between 18 and 80 years of age with LTBI or with pulmonary ATB disease were recruited. Those who had 1) immune deficiency disease; 2) glucocorticoid treatment > 1 week; 3) serious heart, liver, kidney, and spleen and other organ diseases; 4) serious allergies; were excluded from the study

analysis.

3.2.5 Sample Preparation

Plasma samples were prepared using the following protocol. 400 μL solvent of methanol/acetonitrile (1: 1, v/v, containing internal standard 2-chloro-L-phenylalanine, prechilled to -20 oC) was inserted into 100 μL plasma. The sample was then vortexed for 30 s followed by incubating at −20 oC. After 2 hours, the mixture was vortexed again and centrifuged at 12,000 r/min at 4 oC for 15 min. For the hydrophilic interaction liquid chromatography (HILIC) analyses, one aliquot of each sample was prepared in 100 μL acetonitrile/water (1:1, v/v). For the reversed-phase liquid chromatography (RPLC) analyses, the other aliquot was made using 100 μL methanol/water (4:1, v/v). The quality control (Q.C.) samples were prepared by mixing equal volumes (10 μL) of each plasma sample to be a pooled plasma sample. The QC samples were prepared following the same protocol and periodically added for every 10 test samples throughout the analytical run.

3.2.6 Metabolomic Analysis and Data Preprocessing

We follow similar methods by Zhu et. al in conducting LC-MS and LC-MS/MS as well as data pre-processing. A detailed description of the parameters is included in the supplementary material section.

3.2.7 Statistical Analysis

We followed the computational and statistical analysis routine[224] in this study. A total of 191 participants from 4 different groups were included in both univariate and multivariate analysis. 143 of them were set as discovery set and the remaining 48 were validation set. The subject characteristics are shown in Table 3.1. The logistic regression models were used to classify whether the individual is an HC, ATB, NTB, or LTBI using Kyn, Trp, and IDO. The logistic regression classification model uses a logistic function to get the probability of being a certain class between 0 and 1 and predict the binary class and can be further extended to multi-class classification. We chose the logistic regression model because it is very efficient to train and gets good interpretation, which can interpret model coefficients as an indicator of biomarker importance. We first performed univariate analysis to estimate the biomarkers separately and then used them together to diagnose active TB. logistic regression classification models with internal 5-fold cross-validation were first performed on the discovery set. Once the models were trained well on the discovery set, then we validated our classification model on the validation set to evaluate their performance. Finally, we collected an external independent cohort to verify our results. The performance of our models was estimated using accuracy, specificity, sensitivity, the AUC, and the ROC curve with the error bars of 95% confidence intervals (calculated on the discovery set using 5-fold cross-validation). We used the error bars to verify the robustness of our models. All the

model fitting methods presented in this thesis used custom scripts in Python.

3.3 Results

3.3.1 Subject Characteristics

A total of 143 subjects participated was recruited in this study for the discovery set (refer to Table 3.1). There were 37 HC (age, 28 [range 19-40] years; males, n = 33 [89%]), 34 subjects with ATB (age, 47, [range 1-85] years; males, n=24 [71%]), 35 subjects with NTB (age, 60 [range 3-84] years; males, n = 21 [60%]), and 37 subjects with LTBI (age, 18 [range 18-18] years; males, n = 19 [51%]). For the validation set, there were 13 HC (age, 22 [range 18-39] years; males, n = 11 [85%]), 12 subjects with ATB (age, 31.5 [range 10-71] years; males, n=10 [83%]), 11 subjects with NTB (age, 55 [range 4-80] years; males, n = 9 [82%]), and 12 subjects with LTBI (age, 18 [range 18-18] years; males, n = 7 [58%]).

A total of 9822 peaks were detected using the HILIC mode, whereas 7200 peaks were obtained using the RPLC-ESI+ mode and 5501 peaks obtained using the RPLC-ESI- mode. The peaks detected out of the retention time range were removed. The peaks generated by the internal standard were also removed. There were 9811 peaks left for the HILIC mode (expressed as the HLIC data set), whereas 7039 peaks were left for the RPLC-ESI+ mode (expressed as the RP-ESI+ data set) and 5485 peaks left for the RPLC-ESI- mode (expressed as the RP-ESI- data set). These data sets were normalized

and then used for further statistical analysis and machine learning studies. In this study, we used z-score standardization by removing the mean and scaling to unit variance.

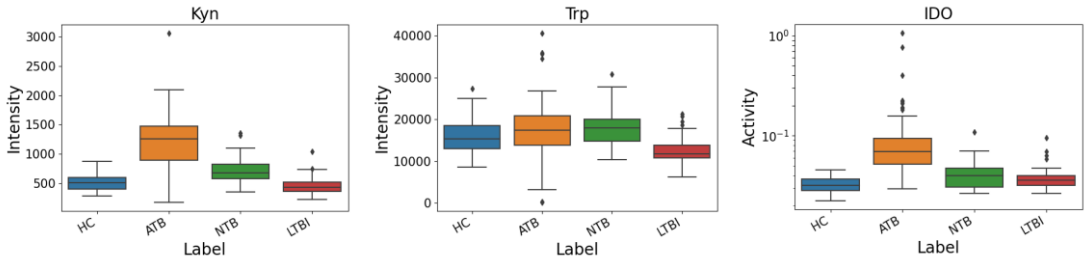
3.3.2 Univariate Analysis

We first employed univariate analysis to check the influence of each metabolic separately. Binary classification models using Kyn, Trp, or IDO, separately were built to distinguish between different groups, HC vs ATB, LTBI vs ATB, and NTB vs ATB. LTBI refers to the presence of Mycobacterium TB in the body, but there are no clinical syndromes of active TB infection. Cause these groups can be infected by Mycobacterium TB for a lifetime without getting sick, we can treat them as a healthy group. So, we combined LTBI and HC into one group as the control group and continually built classification models between the control group vs ATB. We performed two different statistical hypothesis tests (t-test and Mann-Whitney U test) on the distribution of these variables in the two populations. Because the t-test requires the data to follow a normal distribution, but the Mann-Whitney U test doesn't,[225] we used the Mann-Whitney U test as a non-parametric alternative to the t-test in this study. The hypothesis test results are summarized in Table 3.2. It shows that except for Trp, Kyn and IDO are statistically significant between all binary combined groups with a p-value smaller than 0.05. We also generated the box plots with whiskers of these three variables to explore their distributions in different classifications shown in Figure 3.1, indicating that the concentration of Kyn and activity of IDO are significantly higher in

the ATB group than in other groups. The results are in line with our hypothesis tests.

We used accuracy, specificity, sensitivity, and AUC score to evaluate the performance of the binary classification using a single biomarker. 5-fold cross-validation was performed to verify the performance results. The results are shown in Table 3.2 and Figure 3.2. In Figure 3.2, the blue curve indicates the mean ROC curve in the discovery set among five-folds and the green curve is the ROC curve in the validation set. The red regions show the 95% confidence intervals of the ROC curve. Hypothesis test results and box plots are consistent, indicating that except for Trp in HC vs ATB and ATB vs NTB cases, the others are directly and significantly different between the two groups in classification. Therefore, Kyn and IDO can be useful biomarkers used separately to classify HC vs ATB, LTBI vs ATB, and control vs ATB efficiently with AUC above 0.90 on the validation set, while the results on NTB vs ATB are not as good as the other classifications, just with AUC of 0.80 and 0.86 respectively. The classifications using Trp are not good, which is consistent with the hypothesis test, but it can still contribute to our classifications in some cases as the concentration of Trp is significantly different between ATB and LTBI. The results of these independent verification experiments fully proved that the classifier that only uses a single variable can perform well in some cases, and we believe that these results can be greatly promoted when more variables are included. The binary classifications, and hypothesis test results among other groups without ATB were summarized in Figure A.1 and Table A.3 which indicated that these three biomarkers are not excellent indicators for

distinguishing among HC, LTBI, and NTB.



(a) (b) (c)

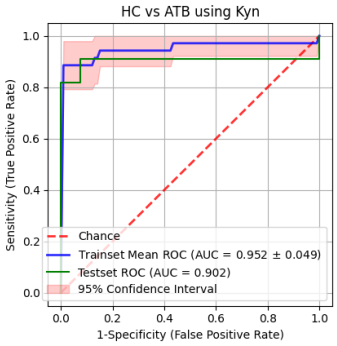
Figure 3. 1 Box and whisker plots for different biomarkers on HC, ATB, NTB, and LTBI patients; (a) Kyn; (b) Trp; (c) IDO. In the box plot, there is a six-number summary of the data, the minimum, first quartile, median, third quartile, maximum, and the outliers. The solid line inside the box represents the median and the whiskers represent the maximum and minimum values, excluding any outliers. The black diamonds outside the whiskers represent the outliers.

Table 3. 2 Performance of logistic regression models with various biomarkers for discriminating different groups along with the hypothesis test results.

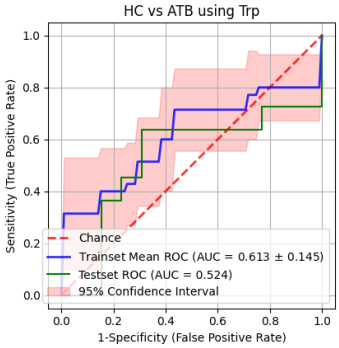
	HC vs ATB					
	Kyn		**Trp**		**IDO**	
	Discovery	**Validation**	**Discovery**	**Validation**	**Discovery**	**Validation**
AUC	0.95 (+/- 0.05)	0.90	0.61 (+/- 0.15)	0.52	0.93 (+/- 0.05)	0.96
Accuracy	0.89 (+/- 0.07)	0.92	0.61 (+/- 0.10)	0.58	0.79 (+/- 0.04)	0.79
Specificity	0.97 (+/- 0.05)	1.00	0.76 (+/- 0.08)	0.69	1.00 (+/- 0.00)	1.00
Sensitivity	0.80 (+/- 0.13)	0.82	0.46 (+/- 0.23)	0.46	0.57 (+/- 0.08)	0.55

	Mann-Whitney *U* test	*t*-test	Mann-Whitney *U* test	*t*-test	Mann-Whitney *U* test	*t*-test
P-value	4.75E-14	1.25E-12	0.088	0.27	2.72E-14	0.0013
	LTBI vs ATB					
	Kyn		**Trp**		**IDO**	
	Discovery	**Validation**	**Discovery**	**Validation**	**Discovery**	**Validation**
AUC	0.96 (+/- 0.05)	0.92	0.73 (+/- 0.11)	0.72	0.85 (+/- 0.08)	0.94
Accuracy	0.90 (+/- 0.03)	0.92	0.73 (+/- 0.10)	0.79	0.73 (+/- 0.08)	0.79
Specificity	0.98 (+/- 0.04)	1.00	0.81 (+/- 0.10)	0.92	0.95 (+/- 0.06)	1.00
Sensitivity	0.82 (+/- 0.05)	0.83	0.64 (+/- 0.17)	0.67	0.50 (+/- 0.13)	0.58
	Mann-Whitney *U* test	***t*-test**	**Mann-Whitney *U* test**	***t*-test**	**Mann-Whitney *U* test**	***t*-test**
P-value	3.70E-14	1.33E-13	7.19E-5	0.00037	6.06E-11	0.0024
	NTB vs ATB					
	Kyn		**Trp**		**IDO**	
	Discovery	**Validation**	**Discovery**	**Validation**	**Discovery**	**Validation**
AUC	0.88 (+/- 0.07)	0.80	0.49 (+/- 0.15)	0.47	0.84 (+/- 0.11)	0.86
Accuracy	0.81 (+/- 0.06)	0.70	0.48 (+/- 0.13)	0.44	0.72 (+/- 0.12)	0.74
Specificity	0.89 (+/- 0.05)	0.92	0.34 (+/- 0.25)	0.25	0.91 (+/- 0.15)	0.92
Sensitivity	0.74 (+/- 0.09)	0.46	0.63 (+/- 0.26)	0.64	0.54 (+/- 0.15)	0.55
	Mann-Whitney *U* test	***t*-test**	**Mann-Whitney *U* test**	***t*-test**	**Mann-Whitney *U* test**	***t*-test**
P-value	9.34E-10	1.89E-8	0.43	0.94	1.10E-8	0.0036
	Control vs ATB					
	Kyn		**Trp**		**IDO**	
	Discovery	**Validation**	**Discovery**	**Validation**	**Discovery**	**Validation**
AUC	0.93 (+/- 0.08)	1.00	0.66 (+/- 0.11)	0.70	0.90 (+/- 0.05)	0.95
Accuracy	0.90 (+/- 0.09)	0.97	0.71 (+/- 0.03)	0.73	0.77 (+/- 0.05)	0.78

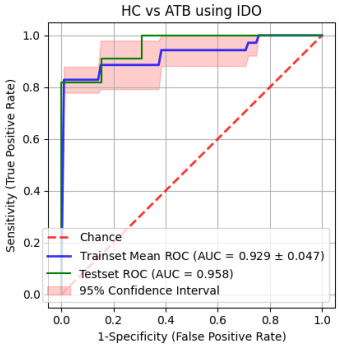
Specificity	0.99 (+/- 0.02)	0.96	0.99 (+/- 0.03)	1.00	0.99 (+/- 0.02)	1.00
Sensitivity	0.71 (+/- 0.22)	1.00	0.12 (+/- 0.05)	0.17	0.30 (+/- 0.15)	0.33
	Mann-Whitney *U* test	***t*-test**	**Mann-Whitney *U* test**	***t*-test**	**Mann-Whitney *U* test**	***t*-test**
P-value	4.59E-18	4.57E-13	0.0016	0.016	4.82E-16	0.0018



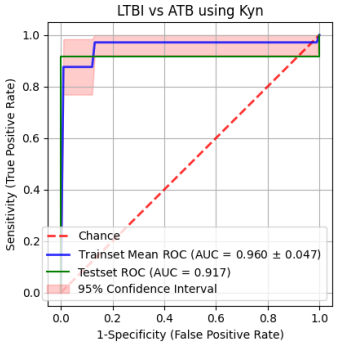
(a)



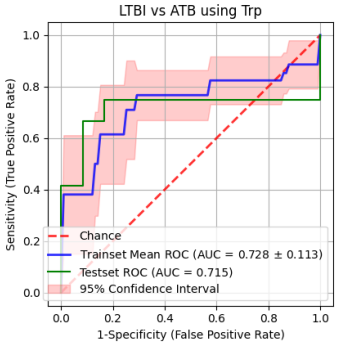
(b)



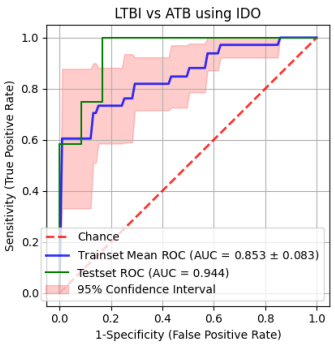
(c)



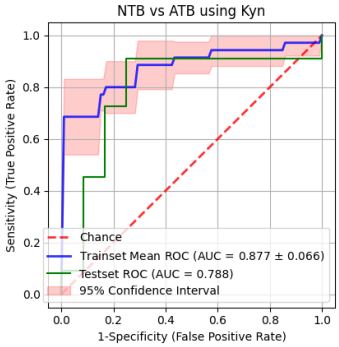
(d)



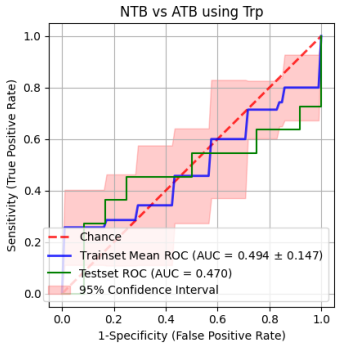
(e)



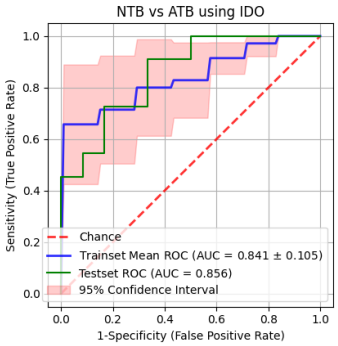
(f)



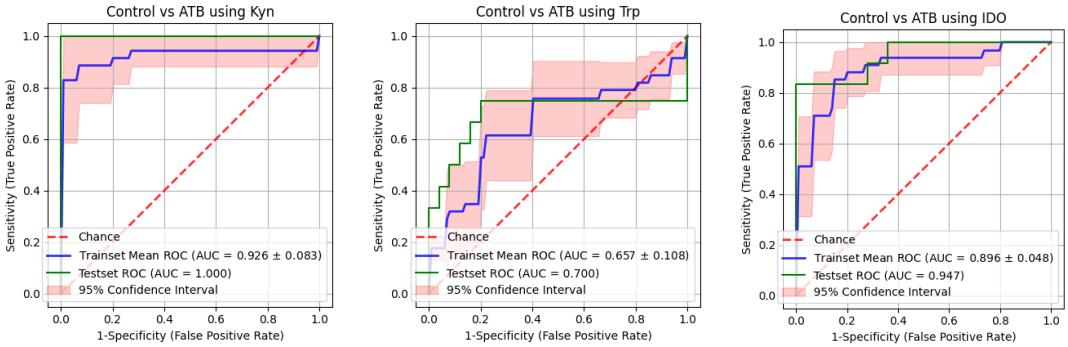
(g)



(h)



(i)



(j) (k) (l)

Figure 3. 2 ROC curves of the logistic regression model: (a) using Kyn for discriminating HC and ATB patients; (b) using Trp for discriminating HC and ATB patients; (c) using IDO for discriminating HC and ATB patients; (d) using Kyn for discriminating LTBI and ATB patients; (e) using Trp for discriminating LTBI and ATB patients; (f) using IDO for discriminating LTBI and ATB patients; (g) using Kyn for discriminating NTB and ATB patients; (h) using Trp for discriminating NTB and ATB patients; (i) using IDO for discriminating NTB and ATB patients; (j) using Kyn for discriminating control and ATB patients; (k) using Trp for discriminating control and ATB patients; (l) using IDO for discriminating control and ATB patients.

3.3.3 Multivariate Analysis

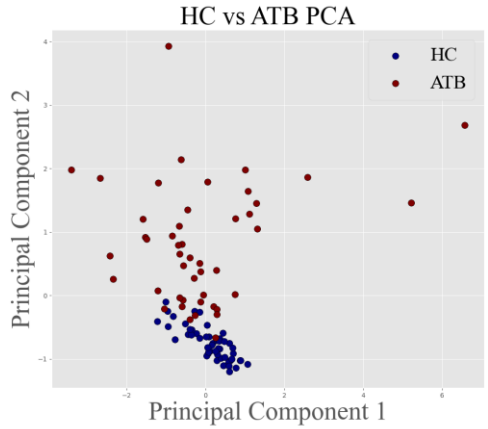
3.3.3.1 Binary Classification Modelling

From the last section, we know that Kyn, Trp, and IDO all can contribute to the classification and perform well separately in some binary groups. To further improve the performance of the classifiers, we included the metabolites together and performed multivariate analysis. PCA with the data from different combined groups were first

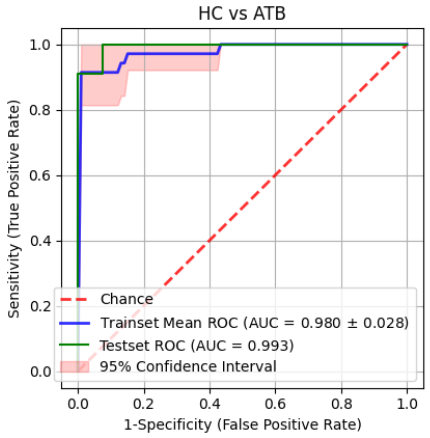
performed and visualized the first two components in Figure 3.3 (a, c, e, g). PCA is an unsupervised multivariate analysis method. In metabolomics, it can reflect the overall difference between samples in different groups. The result is clear and easy to read, and it is widely used in metabolomic studies to explore the differences between different groups. Within the two-dimensional score plot, we found that the patient with ATB could be easily distinguished from other groups, except for the NTB vs ATB, there are small intersections between them. Four binary classification models were employed using Kyn, Trp, and IDO together to classify HC vs ATB, LTBI vs ATB, NTB vs ATB, and control vs ATB. The evaluation metrics including AUC score, accuracy, specificity, and sensitivity are summarized in Table 3.3 and the ROC curves are shown in Figure 3.3 (b, d, f, h). The results indicate that the AUC score on the validation set of these four classifiers used to distinguish between binary groups all increased to varying degrees. Except for the NTB vs ATB classification, we can see that these three biomarkers give out excellent results on the other three classifications and have an accuracy of around 96% and the AUC score can reach 1.00, while the results on the NTB vs ATB classification case are not so good. The accuracy and sensitivity are below 0.80. As we expected, when more variables are included, all the performance metrics indicate higher accuracy, specificity, sensitivity, and AUC score. The binary classification results that distinguish between other groups were summarized in Figure A.2 and Table A.4.

Table 3. 3 Performance of logistic regression models for discriminating different binary groups.

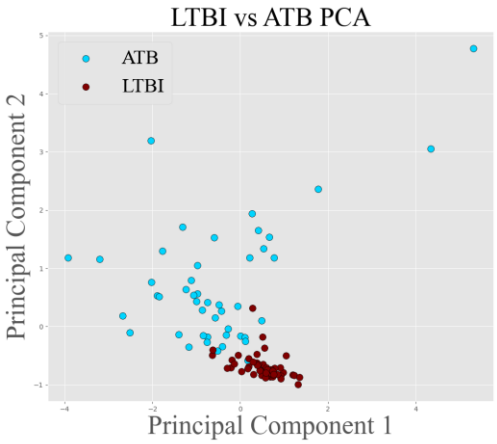
	HC vs ATB		LTBI vs ATB	
	Discovery	**Validation**	**Discovery**	**Validation**
AUC	0.98 (+/- 0.03)	0.99	0.96 (+/- 0.04)	1.00
Accuracy	0.89 (+/- 0.07)	0.96	0.86 (+/- 0.08)	0.96
Specificity	0.97 (+/- 0.05)	1.00	0.92 (+/- 0.06)	1.00
Sensitivity	0.80 (+/- 0.13)	0.91	0.79 (+/- 0.11)	0.92
	NTB vs ATB		**Control vs ATB**	
	Discovery	**Validation**	**Discovery**	**Validation**
AUC	0.88 (+/- 0.08)	0.89	0.97 (+/- 0.05)	1.00
Accuracy	0.80 (+/- 0.07)	0.74	0.93 (+/- 0.07)	0.97
Specificity	0.89 (+/- 0.05)	0.92	0.99 (+/- 0.02)	0.96
Sensitivity	0.71 (+/- 0.11)	0.55	0.80 (+/- 0.17)	1.00

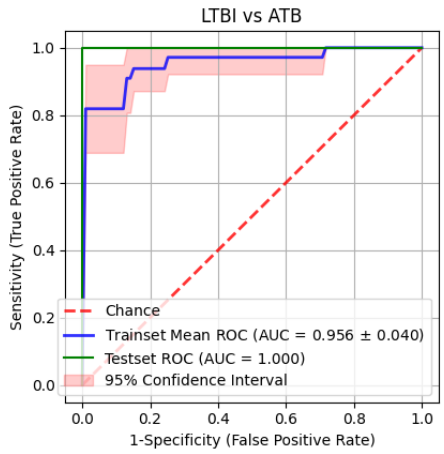


(a)

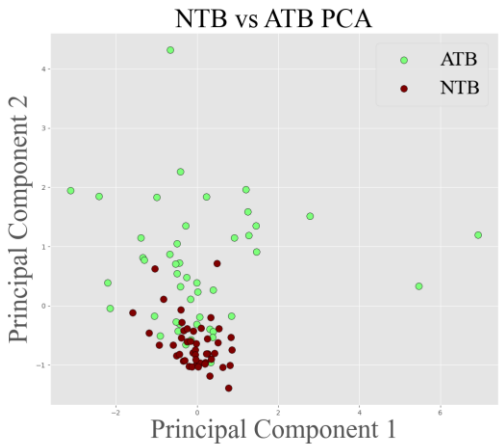


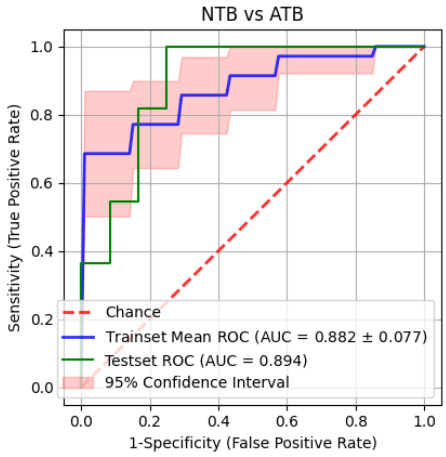
(b)



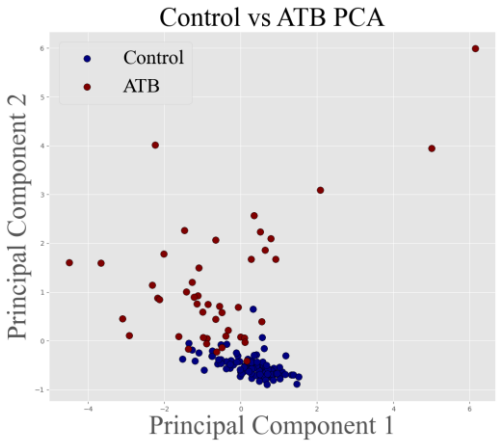


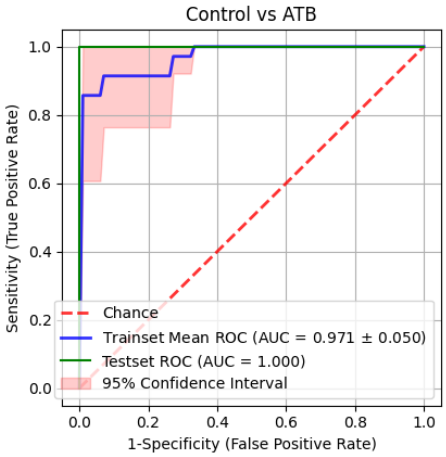
(c) (d)





(e) (f)





(g) (h)

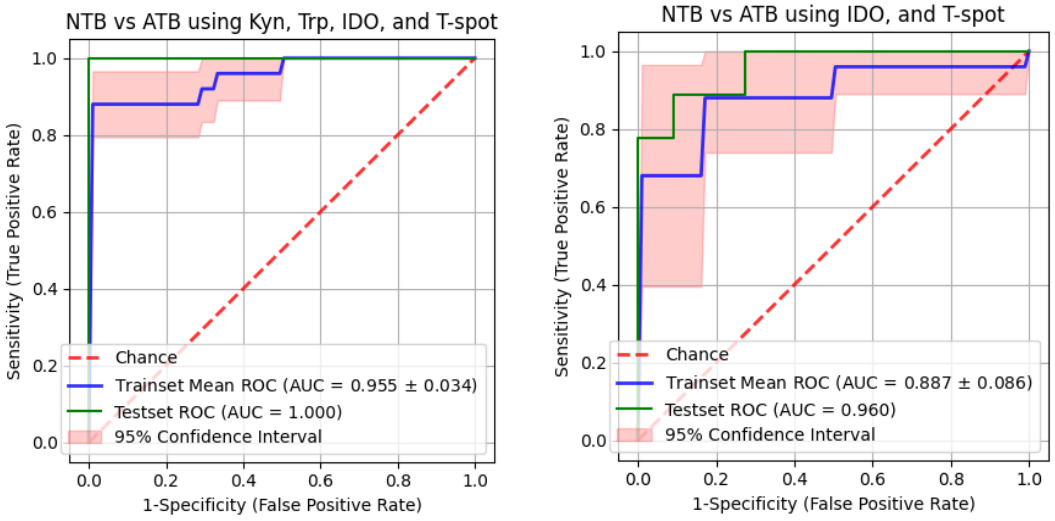
Figure 3. 3 PCA plot shows the ability to discriminate different groups: (a) discriminating HC and ATB patients; (c) discriminating LTBI and ATB patients; (e) discriminating NTB and ATB patients; (g) discriminating control group and ATB patients. ROC curves of the logistic regression model using Kyn, and IDO: (b) discriminating HC and ATB patients; (d) discriminating LTBI and ATB patients; (f) discriminating NTB and ATB patients; (h) discriminating control and ATB patients.

3.3.3.2 Enhanced Binary Classification and External Validation

For now, we have built four binary classifications for multivariate analysis. Three of them got excellent performance with the AUC score of around 1.00 on the validation set, while the NTB vs ATB case cannot give out satisfying results with low accuracy and sensitivity. Here, I proposed a method to enhance the performance of the classifications. To further validate our results, we collected 28 samples from an independent validation cohort as an external validation set to do the double-blind experiments. We were blinded to the clinical diagnoses while performing the statistical analyses.

We included the t-spot result to promote the NTB vs ATB classification. The t-spot test detects and counts the number of effector T cells activated by TB antigen, which can reveal the presence of TB infection. If the t-spot number is equal to or greater than 6, the t-spot result is positive (represented with 1), otherwise, it is negative (represented with 0). Therefore, we got another indicator via the positive/negative t-spot results. First, we removed the samples in NTB and ATB groups for that couldn't collect the t-spot information. The classification results just using t-spot data were summarized in Table A.5 and Figure A.3, showing that the t-spot data itself can help for distinguishing ATB from NTB but cannot yield a satisfying performance. Then we used Kyn, Trp, and IDO along with the t-spot results to distinguish ATB from NTB. After including the t-spot result, although the AUC score on the validation set increased from 0.89 to 1.00, the

performance on the external validation set is still not good, only with 68% accuracy (Figure 3.4, Table 3.4). Therefore, we just used IDO and t-spot to predict NTB and ATB. The accuracy increased from 74% to 80% and got an AUC score of 0.96, which indicates that after including the t-spot result this classification model can be effective to classify ATB and NTB. The accuracy on the external validation set is 82%, which is consistent with our previous results and further proved this classification method is effective and has the potential to be used in real applications.



(a) (b)

Figure 3. 4 ROC curves of the logistic regression model for discriminating NTB and ATB patients: (a) using Kyn, Trp, IDO, and t-spot; (b) using IDO and t-spot.

Table 3. 4 Performance of logistic regression model for discriminating NTB vs ATB.

	Using Kyn, Trp, IDO, and t-spot	**Using IDO, and t-spot**

	Discovery	Validation	External Validation	Discovery	Validation	External Validation
AUC	0.96 (+/- 0.03)	1.00	0.82	0.89 (+/- 0.09)	0.96	0.85
Accuracy	0.88 (+/- 0.04)	0.90	0.68	0.82 (+/- 0.09)	0.80	0.82
Specificity	0.91 (+/- 0.07)	1.00	0.90	0.75 (+/- 0.20)	0.64	0.70
Sensitivity	0.84 (+/- 0.13)	0.78	0.56	0.92 (+/- 0.09)	1.00	0.89

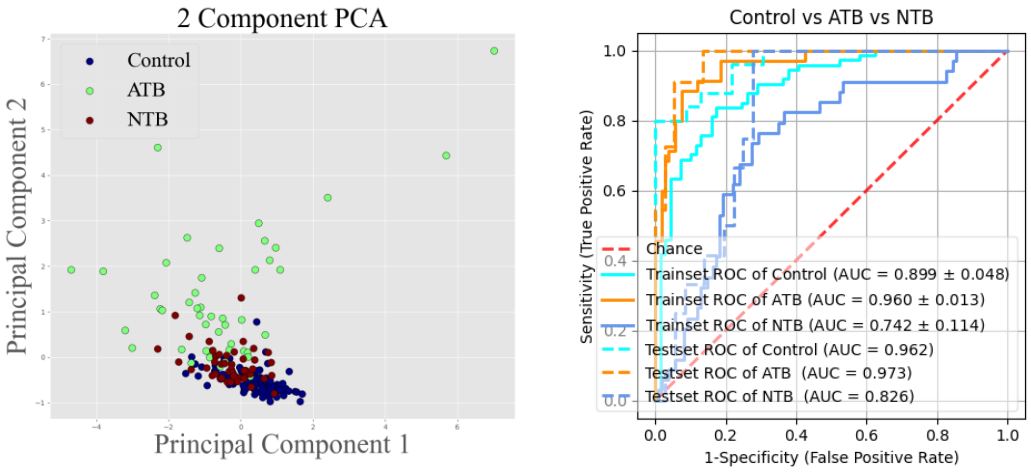
3.3.3.3 Multi-class Classification Modelling

We showed that the Kyn, and IDO can be good indicators to separate the binary categories with high accuracy and AUC score. However, sometimes there are cases where the binary classification cannot meet our requirements, and a multi-class classification model is in need. Here we present the three-class classification model classifying control, ATB, and NTB. We also built the PCA model for three-class classification. Within the two-dimensional PCA score plot (Figure 3.5 a), we found that the cluster of the ATB group can be separated from other samples clearly, while there are small intersections and overlapping among all other groups. The PCA model can effectively identify the ATB but cannot decide whether a participant is a control or NTB. Similar to the binary classification, we performed two different statistical hypothesis tests, one-way ANOVA and the Kruskal-Wallis H test. One-way ANOVA test assumes

that the variables follow a normal distribution, while the Kruskal Wallis test does not[225]. Therefore, we chose the Kruskal Wallis test as a non-parametric instead of the ANOVA test. All Kyn, Trp, and IDO are directly and significantly different in these three populations. The logistic regression model performance is summarized in Table 3.4 and Figure 3.5 b. We can see all the AUC score is above 0.80 and the average AUC score on the validation set is 0.92 and the accuracy is 73% using Kyn, Trp, and IDO. The overall performance is not as good as binary classification. But this three-class logistic regression model has good performance on class ATB, which is consistent with the PCA results. We believe if more variables are included, the multi-class classification model can be gratifying.

Table 3. 5 Performance of logistic regression model for discriminating control, ATB, and NTB using Kyn, Trp, and IDO and hypothesis tests.

Set	**Accuracy**	**Precision**	**Recall**	**AUC**		
				Control	**ATB**	**NTB**
Discovery	0.73 (+/- 0.06)	0.76 (+/- 0.06)	0.73 (+/- 0.06)	0.90 (+/- 0.05)	0.96 (+/- 0.01)	0.74 (+/- 0.11)
Validation	0.73	0.74	0.73	0.96	0.97	0.83
P-value (ANOVA test)			**P-value (Kruskal-Wallis test)**			
Kyn	**Trp**	**IDO**	**Kyn**	**Trp**	**IDO**	
6.27E-27	2.48E-4	4.47E-23	2.30E-07	3.13E-5	3.40E-16	



(a) (b)

Figure 3. 5 (a) PCA plot shows the ability to discriminate among control, ATB, and NTB patients using Kyn, Trp, and IDO. (b) ROC curves of the logistic regression model for discriminating control, ATB, and NTB patients using Kyn, Trp, and IDO.

3.4 Discussion

We reported that together with Kyn, Trp, IDO, and t-spot results can be an efficient way to distinguish ATB from HC, LTBI, and NTB. Accordingly, metabolomics technology has been widely used in various diseases to screen new biomarkers[211]. Published studies have found that IDO, is an immunoregulatory enzyme that regulates the pathogenesis of a variety of pathological conditions, including cancer[226] and infectious diseases[227], [228]. The biochemical role is IDO catalyzes the degradation of Trp to Kyn to inhibit T cell proliferation and function, leading to immunosuppression and tolerance[229]. Therefore, it can be a good indicator for the diagnosis and prognosis of TB[213], [214], [229], [230]. Normally, the activity of IDO, measured by the ratio

of Kyn degraded to Trp concentrations, increases significantly in ATB patients. Adu-Gyamfi et al. reported that the IDO activity in patients with TB was higher than those in healthy control[213]. Suzuki et al. found pulmonary TB patients had a higher concentration of Kyn and significantly lower Trp concentration, resulting in significantly higher IDO activity. However, most of these previous studies focused on distinguishing between ATB and the healthy control, few of them have applied its diagnostic significance in LTBI patients and none of them has conducted experiments on NTB patients. LTBI represented to those individuals infected with MTB remain asymptomatic, despite the ongoing immune response[231]. One-third of the world's population has been affected by the LTBI, and 5%-10% of them will progress to active TB disease in the future according to the WHO TB reports[196]. So, the differential diagnosis between LTBI and ATB is also the key point to control or prevent TB infections. The NTB patients, usually have similar symptoms with ATB patients and are confused sometimes by clinical diagnosis, however, the treatments are different between them. Therefore, it is important in the clinical application to differentiate the ATB from NTB patients. We assessed the diagnostic potential of the concentration of Kyn and Trp, and the activity of IDO, along with the t-spot positive/negative result in a cohort with 4 different groups, and further got verified on an independent cohort.

In this study, we first performed targeted metabolomics on Kyn and Trp and calculated the activity of IDO using their ratio, Kyn/Trp according to the previous studies[213], [214]. Since our study was conducted in a targeted manner, we can

measure these metabolites and their ratio more accurately and reliably than those in an untargeted manner. Machine learning has been widely used and is becoming a key method in both targeted and non-targeted metabolomics[232]. Compared to only using statical analysis, we also integrated machine learning algorithms, logistic regression to establish several effective classification models. The standard logistic regression model predicts the probabilities of an individual being one of the two classes based on a group of metabolites intensities and has been widely used in metabolomics studies. In the univariate analysis, we can see that compared to the patients with LTBI, NTB, and healthy control, the concentration of Kyn and the activity of IDO are much higher in the ATB group. The Mann-Whitney U test gives us consistent results that the concentration of Kyn and the activity of IDO have a statistically significant difference between ATB and other groups. The increase of IDO activity is attributed to the increase in product Kyn and the decrease of substrate Trp concentrations, however, Trp concentration level alone does not have high diagnostic potential. Our findings are in line with the available literature[213], [214], [230] that the IDO activity can distinguish between ATB patients and the HC with the AUC score of 0.96. IDO can also determine an individual with LTBI or ATB with a high AUC score, however, it cannot perform well on distinguishing ATB from NTB alone and it shows no difference among HC, LTBI, and NTB groups, nor diagnostic potential. In addition, the concentration of Kyn alone can also differentiate patients with ATB from HC and LTBI with an AUC score above 0.90.

Univariate analysis results indicated that both the concentration of Kyn and the activity of IDO classification models can reach an AUC score of 0.90 for classifying HC vs ATB and LTBI vs ATB, however, the sensitivities are only 55% and 58%, respectively. Compared to utilizing these biomarker candidates separately, the overall performance increased a lot after we combined the biomarker candidates and performed the classification experiments. We determined diagnostic AUC score, accuracy, sensitivity, and specificity. The classification models have a specificity and sensitivity of 100%, 91%, 100%, and 92%, respectively on distinguishing ATB from HC, and ATB from LTBI. It performed excellently for classifying ATB from both HC and LTBI. After we merged the healthy control and the patients with LTBI, and just distinguished between the control group and ATB, the accuracy, specificity, and sensitivity can reach 97%, 96%, and 100%, respectively. At present, we have obtained three excellent classifiers with satisfying specificity and sensitivity distinguishing ATB from HC, LTBI, and the merged control group. While, as the Kyn, Trp, and IDO cannot satisfy our requirements on distinguishing between NTB and ATB with an accuracy of 74% and a sensitivity of 55%, I proposed an enhanced classification model with the help of the t-spot. T-spot test is a simplified enzyme-linked immunospot assay, designed to detect effector T cells that respond to the specific antigen stimulation of MTB and can get a high specificity theoretically[233]. The t-spot data alone can only achieve a specificity of 64% and a sensitivity of 100%. After including t-spot data, our classification model yielded a good result with the accuracy, specificity, and sensitivity of 90%, 100%, and

78%, respectively. Our results suggest that the concentration of Kyn and Trp, IDO activity, and the t-spot data can be efficient for indicating the presence or absence of ATB disease.

We further validated the ability of these biomarkers to accurately distinguish ATB from HC, LTBI, and NTB in an independent cohort. Even the dataset is obtained from other periods, our external validation experiment (double-blind experiment) can achieve high accuracy of 82% and sensitivity of 89% to classify NTB and ATB. These experiments can sufficiently indicate that our model can distinguish ATB from NTB efficiently with better robustness and stability and has the potential to be developed into an accurate ATB diagnosis device. Last but not least, I proposed a multi-class classification model using Kyn, Trp, and IDO and got an accuracy of 73% on the validation set. The overall performance for the multi-class model is not good, but this classification model can perform well especially on the ATB class with an AUC score of 0.97, which means it can diagnose ATB from control and NTB efficiently. If more biomarkers are included, the results of this multi-class can be proved.

The current study has its particular advantages. We got sufficient data and our samples came from four different groups, HC, LTBI, NTB, and ATB. The classification models I proposed can not only efficiently distinguish ATB from HC and LTBI but also classify NTB and ATB with satisfying performance. Further, our classification results have been validated on an independent cohort, where the researchers were blinded to the clinical diagnosis until all data analysis was completed. Thus, our study has

identified three biomarkers along with t-spot data that provide discriminatory capacity for ATB disease. The limitation in this study is that we only measured three metabolic biomarkers according to the previous studies, and we believe that if more biomarkers are included, the performance of the multi-class classification and the classifications among HC, LTBI, and NTB can also be satisfied.

Overall, our models utilizing these indicators such as Kyn, Trp, IDO, and t-spot have excellent validity in diagnosing ATB from NTB, HC, and LTBI efficiently. A diagnosis with higher accuracy can be expected through these processes, which facilitates us to provide more targeted and timely treatment for tuberculosis. Validated biomarkers of MTB can be used to diagnose and prognose the progress of active tuberculosis and potentially monitor anti-tuberculosis treatment and can make significant progress in the fight against TB.

3.5 Conclusions

Current research shows that the IDO activity in TB patients was predominantly higher than that of other subjects. In conclusion, we found that Kyn, Trp, and IDO are significantly different among these groups. We built several high-performance logistic regression classification models for the diagnosis of ATB. The binary classifications can achieve AUC scores over 0.96 on validation sets. The AUC performance of multi-class classifications is generally greater than 0.83. For the ATB class, these classifiers can always keep the AUC scores above 0.95. In a real-life application, the classifier that

distinguishes ATB from HC and LTBI cannot meet the requirements, we also want to figure out whether the patients get NTB. Our study shows that the classifiers using IDO, along with the t-spot are effective in discriminating ATB, and NTB and have been verified in double-blind experiments. We conducted this study to propose and test Kyn, Trp, and IDO activity as novel biomarker indicators for the detection of ATB with the help of the t-spot. This study is only a pilot study, and non-targeted metabolomic is needed to be performed to add more significant biomarkers to enhance the multi-class classification. This study can contribute to developing diagnostic procedures in combination with other biomarkers and can be easily transferred into a high-performance, low-cost, noninvasive, and rapid pulmonary TB diagnosis and prognosis device.

4 Ultrasound-assisted magnetic nanoparticle-based gene delivery

4.1 Introduction

Gene delivery is now a popular research area with high demand on the market, and applications in both clinical and scientific biomedical research[103], [104]. The applications include but are not limited to, treating cancers, immune-deficient diseases, and genetic diseases[108]. Mammalian cells have a selectively permeable plasma membrane that protects them from the external environment. Effective methods to transfect cells are needed. For the delivery of genetic material into the nucleus of the cell, two approaches can be suggested: increasing the cell membrane permeability and thus facilitating the penetration of the target gene or developing a carrier that can go through the cell membrane, carry the gene, and deliver it to the nucleus. Based on these two different pathways, gene delivery utilizes either chemical or physical methods[106], [107]. The chemical approaches can be further divided into viral and non-viral approaches[106]. The ideal carrier should be low cost, with high loading capacity, high stability, no or low toxicity, and easy to use[105]. The viral-vector system approach is the most common and widely used method[108], which can achieve very high transfection efficiency. However, the safety concerns related to immunogenicity and the high cost remain the main limitations[107]. Non-viral methods include liposome-based methods[109], calcium phosphate precipitation[110], cationic polymers[111],

[112] (such as polyamidoamine dendrimers and PEI[113]), and nanoparticle-based hybrids[114]. The cationic liposomes are the most commonly used non-viral delivery system for gene delivery. They can reach most of the requirements of the ideal characteristics with the significant drawbacks of high toxicity and inflammatory responses[109]. Calcium phosphate precipitation and PEI get low transfection efficiency and high cytotoxicity[110]. Nanoparticles are submicron-sized polymeric particles, due to the sub-cellular and sub-micron size range, they can penetrate tissues more efficiently[115]. MNP is one of the traditional nanoparticles and is also a popular carrier for gene delivery[116]. MNP can overcome the weaknesses of other traditional carriers, like high toxicity limiting the traditional carriers that can only be used *in vitro*[117]. The external magnetic fields applied on the target site not only can enhance the transfection, but also target the gene to a specific site without the side effects on other tissues. Due to this, MNPs can be tunable and focus on the target area, yet they still have some drawbacks like low transfection efficiency and toxicity[118].

Besides the chemical approach, the physical delivery methods are attracting more and more research interest, including the application of the electric field[234], the acoustic method[120], and physical injection[121], to disrupt the cell membrane and let the DNA pass through it more efficiently. Some physical techniques have shown sufficient delivery efficiency and can be applied to most of the cell types and are available for commercial use[122]. However, the main limitations are cytotoxicity and the inability to be used in humans. Also, operational and equipment requirements are

complicated and costly, and sometimes with low efficiency and repeatability. Acoustic methods are another physical approach to transfect the cells with the advantage of easy repeatability and excellent stability, yet with low transfection efficiency compared with other methods[120].

The ultrasound method is one of the acoustic transfection methods mentioned above, which is characterized by frequencies higher than 20 kHz. Ultrasound has been used for various applications, including diagnosis, surgery, and therapy for a long time[117], [128]. At its early implementations, researchers focused on the treatment produced by using the thermal effects of ultrasound. Nowadays, more researchers are paying attention to the non-thermal characteristic, including acoustic cavitation and mass transfer enhancement[131]. For ultrasound medical applications, the safe range of the intensity is between 0.05 W/cm^2 and 100 W/cm^2[132], [133]. LIPUS is a particular type of ultrasounds shown in Fig 4.1, which generates at a frequency of 1-3 MHz and repeats at 1 kHz with a duty cycle of 20% to deliver a low-intensity pulsed-wave [137]. It has been proven to be very safe for human use for many aspects of medical applications, such as bone healing[134], inflammation inhibiting[135], soft-tissue regeneration[136], and the induction of cell-membrane porosity. Using graphene aerogel to promote cell proliferation was reported[235], [236]. LIPUS has also been proven to help division and proliferation in many types of cells, such as insect cells[138], algal cells[139], stem/progenitor cells [140], mesenchymal stromal cells[237]. LIPUS can also increase CHO cell growth and antibody production[142], increase cell

permeability[138], and enhance gene delivery by using microbubble[143]. The safe operational intensity range of LIPUS is between 0.02 and 1 W/cm^2 and treatment durations of 5 - 20 minutes per day. Because of its low intensity, LIPUS has almost no thermal effects[144].

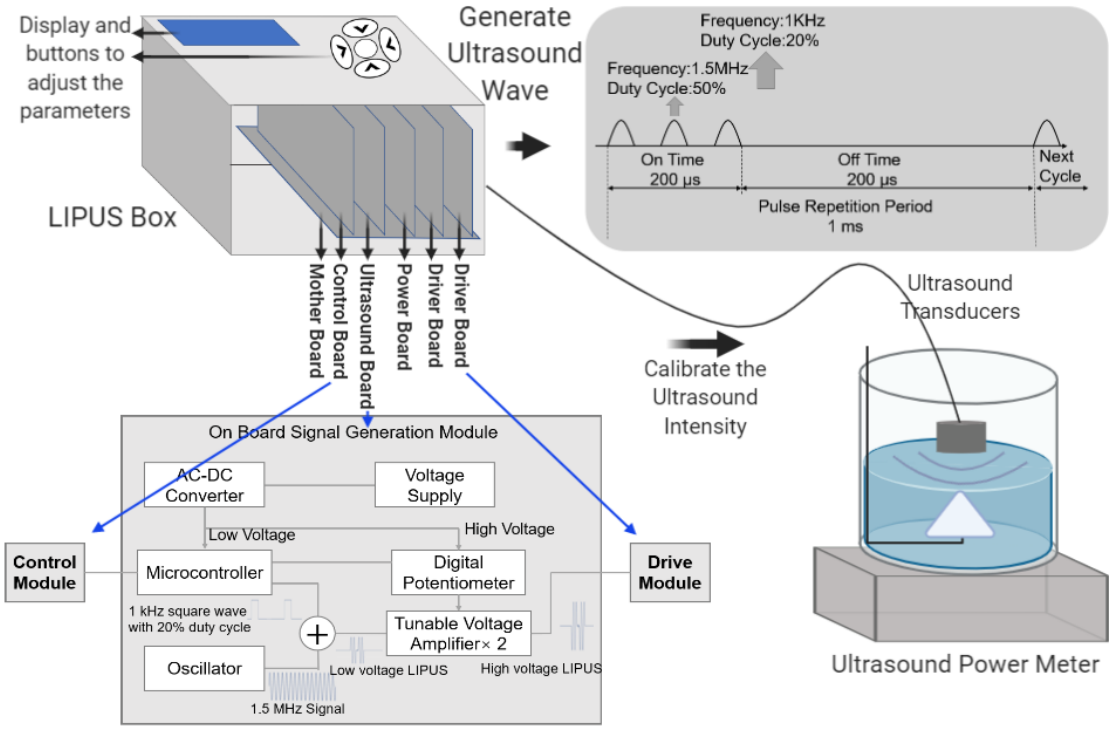


Figure 4. 1 A schematic for LIPUS device and ultrasound power meter calibration. The display shows the ultrasound intensity. The button can be employed to control the duty cycle and ultrasound stimulation duration. The ultrasound boxes include a motherboard, a control board, an ultrasound board, two driver boards, and a power board. We also include the circuit diagram.

There are several ways to deliver genetic materials into cells. As mentioned in the

introduction, I propose to synergistically combine LIPUS and magnetic fields for gene delivery, because the technique can leverage each method's advantages. LIPUS can promote cell growth[141] and enhance cell membrane permeability. At the same time, MNPs, guided by a magnetic field, can further improve targeted gene delivery. Best to our knowledge, combining the two techniques to deliver genetic material into mammalian cells has not been discussed before. We achieved a high transfection efficiency with low cytotoxicity by performing the cell transfection with LIPUS and MNPs under the external magnetic field. We present a new concept of integrating the physical and chemical gene delivery approaches by introducing LIPUS to support gene transfection using MNPs under the influence of the magnetic field. The plasmid of interest (GFP and Cherry Red plasmid) is firstly bound to the MNPs through PEI before introducing them to the cells. Then we investigate the impact of applying the magnetic field in combination with the LIPUS on the transfection of the targeted plasmid. We also examine the effect of the LIPUS on cell proliferation and viability to identify the proper ultrasound intensity and best duration of treatment that cells can tolerate. We used the fluorescent microscope for qualitative evaluation, whether this approach worked or not, and then, employed flow cytometry to quantitatively evaluate the transfection efficiency and compare the results with those obtained from using Lipofectamine 2000 as a positive control. Furthermore, we also localized the transfected genes in the targeted cells using confocal microscopy. We have compared the transfection rates of various gene delivery methods. Please refer to the following

Table 4.1. Our approach offers a combination of high efficiency with low toxicity and affordability.

Table 4. 1 The transfection rate and cell viability of different delivery methods with HEK 293 cells.

Delivery Method	Transfection Rate	Cell Viability
PLGA-PEG/PBAE/pGFP Nanoparticle[238]	45.2%	97%
Magnetoplexes[239]	28.4%	97%
Lipoplexes[239]	45.5%	95%
Lipofectamine 2000	42.6%	44.7%
Thiolated trimethyl amino benzyl chitosan[240]	40.4%	90.0%
Magnetic Nanoparticles[118]	56.1%	58.3%
Our approach (MNPs & LIPUS)	61.5%	63.6%

4.2 Materials and Methods

4.2.1 Chemicals and Materials

Anhydrous Ethylene glycol 99.8%, Ferric chloride hexahydrate (FeCl2.6H2O) ≥ 99 %, and anhydrous Sodium acetate ≥ 99% (NaAc), and branched PEI with average molecular weight (M.W.) 25 kDa were purchased from Sigma-Aldrich and used without further purification. HEK 293T cells were purchased from ATCC (ATCC CRL-11268), MEM, FCS, Penicillin/Streptomycin, PBS, Lipofectamine 2000, and DAPI were purchased from Thermofisher. Zombie Aqua was purchased from Biolegend. We used

Milli-Q water with the resistivity of 18.2 MΩ versus the Millipore Milli-Q Advantage A10 purification system in all experiments.

Magnet: The magnet, made of neodymium (rare earth) with a diameter of 10 cm and a thickness of 10 cm, was purchased from Applied Magnets (Plano, TX, USA).

4.2.2 Cell Culture

HEK 293T cells were cultured in high glucose MEM medium, supplemented with 10% FCS and 1% penicillin/streptomycin at 37°C and 5% CO2. Cells were passed to a 12-well plate before the experiment, and transfection was performed in the 12-well plate once the cells reached 60-80%, confluency, which is recommended for the transfection.

4.2.3 Synthesis and Functionalization of MNPs

MNPs were synthesized using the hydrothermal method, according to our previously reported work[117], [118]. In short, a reaction mixture containing 10 g 1,6-hexanediamide, 2.0 g $FeCl_3$•$6H_2O$, and 4.0 g sodium acetate trihydrate (NaAc •$3H_2O$) in 50 mL of ethylene glycol was vigorously stirred at 85 ºC for 2 h until a resulting transparent solution was obtained. To complete the reaction, the solution was sealed in a 100 mL-Teflon-lined stainless-steel autoclave and put in the oven for 12 h at 200 ºC. After completion, the MNPs solution was cooled down to room temperature and collected with the help of a magnet and further redispersed in milli-Q water by

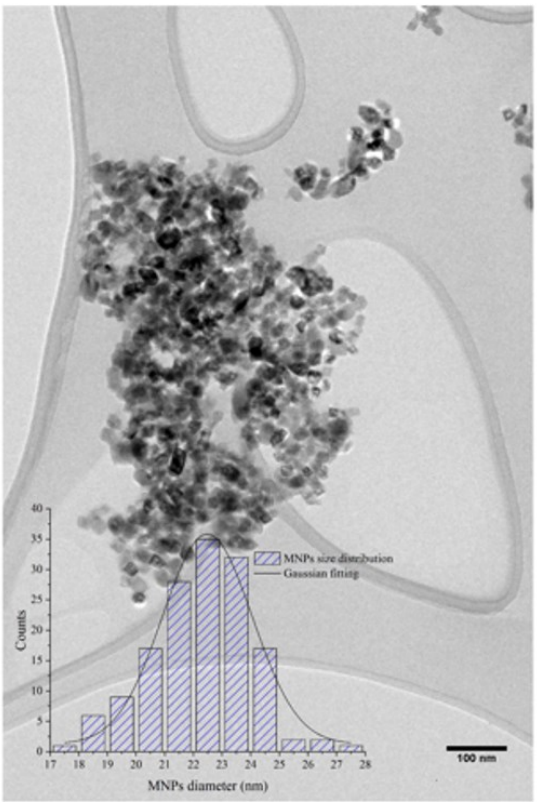
sonication for 15 min. MNPs were washed with water three times, where the MNPs were redispersed by sonication and collected each time with the help of the magnet. Then we also washed with absolute ethanol following the same method to ensure the complete removal of unreacted materials with the abbreviation of FN. Finally, the prepared MNPs (or FN-MNPs) were dispersed in 100 mL milli-Q water for characterization and further use.

For functionalization, FN-MNPs were treated with a 5% glutaraldehyde solution for 2 hours, then washed three times with Milli-Q water and further coated with 1 mg/mL solution of PEI (25 KDa) to produce FN-Glu and FN-Glu-PEI25K, respectively. PEI is known to assist in cell transfection, yet it is toxic to the cells. We selected PEI as a cationic surfactant to coat the MNPs because of its high affinity to bind with negatively charged plasmids. The PEI molecules are covalently bound to the MNPs, which significantly decreases the potential toxicity associated with using cationic surfactants. Furthermore, the complex of the negatively charged plasmid with FN-Glu-PEI25K MNPs mitigates the positive charge of the FN-Glu-PEI25K MNPs. We previously utilized FN-Glu-PEI25K MNPs as gene nano-carriers, and the GFP plasmid/ FN-Glu-PEI25K MNPs showed lower toxicity than the positive control (lipofectamine 2000)[118], [241].

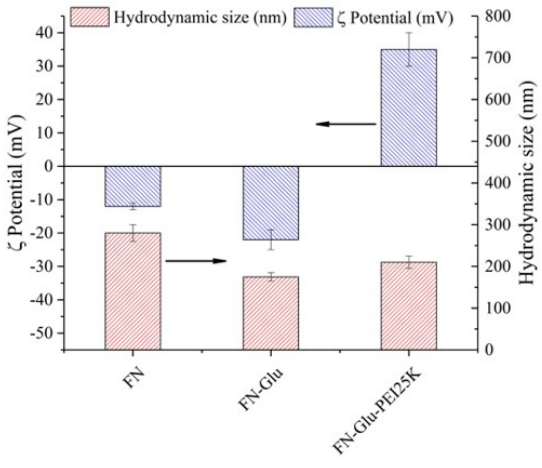
The size of nanoparticles was estimated using Hitachi HF-3300 Transmission Electronic Microscope (300 kVTEM), where we measured the size of 150 nanoparticles using ImageJ software, and the size of the nanoparticles was calculated based on the

histogram of the size distribution. The prepared MNPs were ~24 nm in size, as shown in Fig 4.2 a.

The ζ potential and the hydrodynamic size of the MNPs were measured using Zetasizer Nano ZS Malvern Panalytical. The hydrodynamic size and zeta potential for our nanoparticles were evaluated and shown in Fig 4.2 b. Unlike the FN and FN-Glu, we got a high positive surface charge of Zeta potential for FN-Glu-PEI25K. They all have different hydrodynamic sizes, but FN-Glu-PEI25K showed a higher density of PEI on the MNPs surface for DNA binding.



(a)



(b)

Figure 4. 2 Characterization and functionalization of MNPs (a) MNPs size

distribution under TEM. (b) Hydrodynamic size and ζ potential for particles. Here FN stands for MNPs, FN-Glu stands for MNPs after glutaraldehyde treatment, and FN-Glu-PEI25K stands for MNPs after glutaraldehyde treatment coated with PEI.

4.2.4 Ultrasound Stimulation Device

Cells were exposed to ultrasound stimulation using the LIPUS device developed previously in our lab. The LIPUS device outputs a square wave with a frequency of 1.5 MHz. The repetition rate is 1 kHz, and the duty cycle is 20%. We can adjust the output voltage from 1.25V to 12.5V by adjusting the potentiometer. Along with the increase of the ultrasound intensity, the output voltage increases too. The PCB board is shown in Fig 4.1 and has 6 boards in the box, including a motherboard, a control board, a power board, an ultrasound board, and two driver boards. The motherboard is used to connect all other boards. The control board is used to control the ultrasound intensity and duration. The ultrasound board is used to supply an ultrasound signal with a frequency of 1.5 MHz and a repetition rate of 1 kHz, and the driver board is used to provide enough voltage and current to drive the transducers. Ultrasound settings can be controlled and adjusted.

Ultrasound transducers: Two ultrasound transducers we used in the LIPUS device were purchased from American Piezo Company International, Ltd (Mackey Ville, USA). The piezo-crystal 880 inside the transducer has a diameter of 25 mm and a thickness of 12.5 mm. Although the diameter of each well in a 12-well cell culture plate

is 22 mm, the diameter of the transducer is slightly bigger than the well. However, it does not affect the results because we purposely calibrate the transducer with an intensity of 30mW/cm^2, not the overall power. It has a resonant frequency of 1.5 MHz and a piezoelectric charge constant d33 of 215 m/V. The ultrasound power meter we used to measure the ultrasound intensity was purchased from Ohmic Instruments Co., Maryland, USA, and the model is UPM-DT-1AV. The diameter of the transducer is 25 mm, and thus its active area is 4.9 cm^2. For instance, if we want to have the output intensity of 30 mW/cm2, the measured output power should be 4.9×30=147 mW=0.147 W. The minimum measurement changing the value of this power meter is 0.002 W. Therefore, in a real operation, we adjust the resistance of the potentiometer to get the readings as 0.146 W, or 0.148 W. Each of the transducers was calibrated before the experiment using a degassed water tank, in which transducer was fixed using a holder until the reading of the output was stable.

4.2.5 Cell Counting

Cells were trypsinized and collected in a clean conical tube. We prepared a trypan blue-stained dilution of cells to count. We conducted a 1:2 dilution (50 µL cells + 10 µL Trypan blue + 40 µL PBS). 10 µL of the diluted cells were loaded onto the hemacytometer. We put the micropipette tip into the groove on the slide and touch it gently to the coverslip before we eject; the liquid should flow into the counting chamber by capillary action. The cells need 30 seconds to settle. Finally, we count four quadrants

and calculate the average to get the cell concentration.

4.2.6 Cell Transfection

Transfection was performed at 60-80% cell confluency. GFP plasmid or Red Cherry plasmid were used as genetic material to transfect. 1ug purified plasmid was mixed with 3 µL MNPs both previously diluted in 500 µL serum-free medium to a total volume of 1000 µL and then thoroughly mixed and left for 30 min. The serum can interfere with the formation of the complex of DNA with MNPs[242]. After 30 minutes, the medium was removed from the cell wells and replaced with a mixture of MNPs-DNA complexes. To direct the MNPs into the cells, we then put the plate on the LIPUS/magnet device and treated cells for 10 minutes. After that, we stopped the LIPUS and incubated the cells on the magnet for 4 more hours. The medium was then replaced with a 10% serum MEM medium, and cells were further grown in the incubator for up to 48 hours. Transfection efficiency was checked within 24-48 hours after the experiment.

For the Lipofectamine 2000 transfection, 1 µg purified plasmid was mixed with 3 µL lipofectamine in serum-free MEM and left for 30 mins. We added the DNA-lipofectamine complex to the cells after the 30-minute incubation.

4.2.7 Transfection Evaluation/Characterization

Cell transfection efficiency was evaluated within 48 hours after the experiment

using several methods. Confocal microscopy images were obtained using the Zeiss LSM 710 confocal microscope. For the confocal microscope imaging, cells were cultured and transfected on the coverslip using the same protocol mentioned above, and after 48 hours were fixed in the 4% PFA and stained with DAPI. Fluorescent microscope images were obtained by Zeiss, Axiovert 200 fluorescent microscope. For the qualitative evaluation of transfection during the method development, fluorescent microscopy was used. Cells after transfection were checked within 24-48 hours for the fluorescent protein expression. As a negative control, untreated HEK 293T cells were used. Both groups of cells were fixed in 4% freshly prepared PFA for 10 minutes.

4.2.8 Flow Cytometry

The quantification of the transfection efficiency was performed using Attune X Flow Cytometer. Flow cytometry was performed with Zombie Aqua as our cell viability dye in dilution of 1:250, which was selected based on the pre-experimental titration and gave the clearest separation of the dead and alive cells. In the Flow Cytometry experiment, aside from experimental groups of samples, we used compensation controls, positive control, and negative control to assure the accuracy of the results.

(a) For the negative control, untreated cells without any genetic material and viability dye were used.

(b) As there were two main fluorophores, we used two compensation controls to prevent the spills: For Zombie Aqua compensation, we used cells, stained with viability

dye; For GFP or Cherry Red compensation, we used cells transfected with a plasmid (GFP or Cherry respectively) and Lipofectamine as transfection agent.

(c) For the positive control, the purified plasmid was used for transfection, and Lipofectamine as a golden standard for transfection available on the market, and cells were stained with Zombie Aqua after 48 hours.

(d) For the experimental group, two groups were set to evaluate the effect of LIPUS on the MNPs transfection and cell viability. First group: cells, plasmid, and MNPs were treated with a magnet for 4 hours and stained with viability dye after 48 hours. The other group was additionally treated with a LIPUS device (10 mins duration at 30mW/cm^2 intensity) and a magnet for 4 hours and then stained after 48 hours.

4.2.9 Statistical Analysis

Each experiment was repeated at least three times. The data was presented by including means and standard deviation. The statistical analyses between different groups were conducted by one-way ANOVA paired t-test. $p < 0.05$ were considered as statically significant.

4.3 Results and Discussions

In this thesis, we combined both physical and chemical approaches to develop a high-efficient gene delivery method with low cytotoxicity. From the previous work done in our group as well as by other groups and reported in the literature, we knew

that LIPUS could transiently increase cell membrane permeability[138] as well as can be beneficial for cell viability[141], [142]. Our goal was to evaluate whether ultrasound can enhance the entry of genes into the cells when used with MNPs as a transfection tool.

4.3.1 Selecting Optimal Ultrasound Condition

Different ultrasound parameters, such as wave intensity, treatment duration, and frequency of the treatment, can have a significantly different effect on cell growth and membrane permeability[138], [142]. Therefore, we needed to select the optimal conditions of the LIPUS stimulation for our studies. In our experiments, five different ultrasound intensities and durations were selected. These conditions were selected based on our previous successful studies; for instance, 30 mW/cm^2 showed the optimal performance for stimulating mammalian cells[141] and the 5-20 minutes treatment duration is in a safe range. These conditions were (1) the control (no LIPUS stimulation); (2) LIPUS at 30 mW/cm^2 for 5 minutes; (3) LIPUS at 40 mW/cm^2 for 5 minutes; (4) 30 mW/cm^2 for 10 minutes; and (5) 40 mW/cm^2 for 10 minutes. For the successful penetration of ultrasound waves to stimulate the targeted cells, ultrasound gel has to be applied on the surface of the ultrasound transducer, because ultrasound does not propagate through the air. If we do not use the ultrasound gel, the transmission coefficient of sound intensity can be 0.00923%. When we use the ultrasound gel, the transmission coefficient of sound intensity can be 31.1%, and thus the ultrasound gel is

necessary.

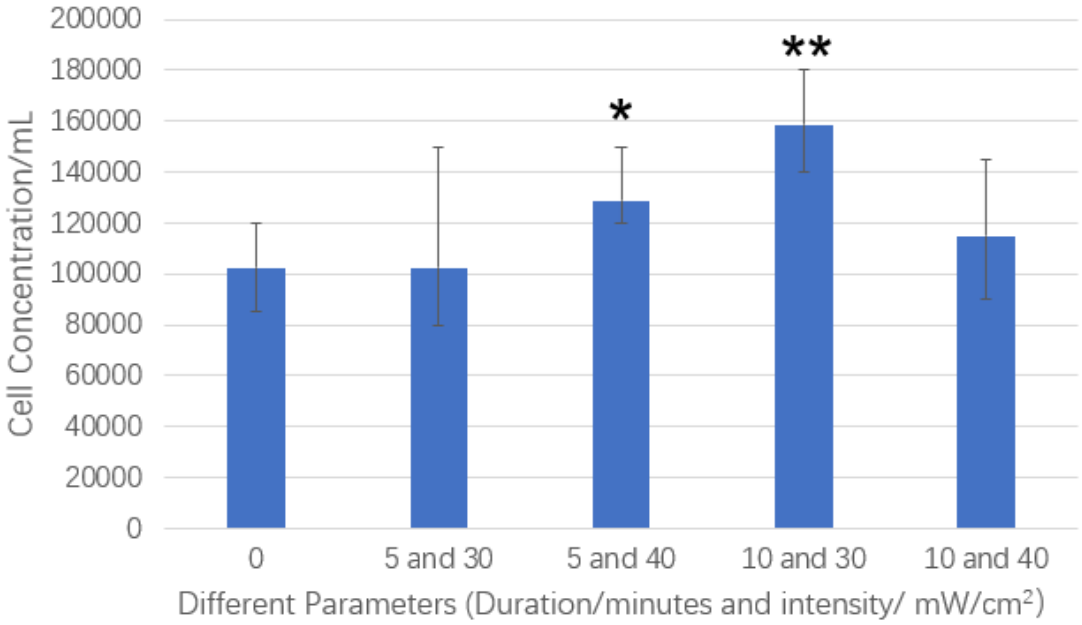


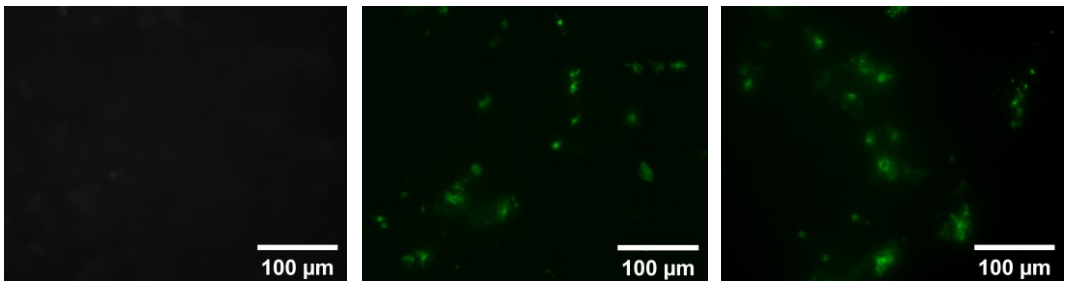
Figure 4. 3 Cell proliferation after stimulation with LIPUS under different intensity and duration parameters. (*:$p<0.05$, **: $p<0.01$).

To select the optimal duration of LIPUS treatment, several stimulation conditions were evaluated, and cell counting was also performed. The results are shown in Fig 4.3, where the cells treated for 10 minutes under 30mW/cm^2 showed the best result with a p-value of 0.00128. P-values for the other stimulation conditions are greater than 0.05 (5 minutes and 30 mW/cm^2: $p=0.5$; 10 minutes and 40 mW/cm^2: $p=0.204$), which means there is no statistical difference Consequently, we use the LIPUS device for 10 minutes under the intensity of 30 mW/cm^2 in our transfection experiments. In our tests, we used two ultrasound transducers at the same time to treat the cells in 2 wells (our

experiment design and setup can allow up to six wells to be treated at the same time).

4.3.2 Fluorescent Microscope Results

Fluorescent microscopy allowed for an easy way to qualitatively evaluate the efficacy of the combined gene delivery method in the process of method development. The negative control, which just contained the GFP plasmid in the cell plates, showed no transfected cells (Fig 4.4 a), confirming that the plasmid itself does not cross the cell membrane without a delivery carrier. Compared to the negative control, using MNPs along with LIPUS treatment introduced the green, fluorescent spots in the images (Fig 4.4 b and 4.4 c), showing the cells that have been successfully transfected with the GFP plasmid and GFP expressed. These fluorescent images were good indicators that our method could work well and were very useful for method development as those allowed for fast qualitative screening of each experiment.



(a) (b) (c)

Figure 4. 4 Fluorescence microscope images: (a) negative control (just cells), (b) (c) cells transfected with GFP with MNPs and treated with LIPUS. Scale bars=100 µm.

4.3.3 Transfection Efficiency using Flow Cytometry

At the same time, the fluorescent microscope images could only give us the qualitative evaluation of whether our new physical and chemical combined approach worked for gene delivery. Once we had our method working, we had to quantify its efficiency and, whether it could be offered as an alternative to the available transfection tools on the market. We were focused on the parameters of transfection efficiency, cell viability, and how they can be compared with the results of the standard gene delivery approach using Lipofectamine 2000, a well-known and efficient transfection reagent, shown in Fig 4.5 b. The average transfection efficiency of the Lipofectamine 2000 in our experiments was 42.62%, as shown in Fig 4.5 a, and it was within the range of normal performance of Lipofectamine 2000 working on HEK 293T cells. This result ensured that our HEK cells were always in good condition before we performed the transfection steps, and we got the proper operations during the transfection.

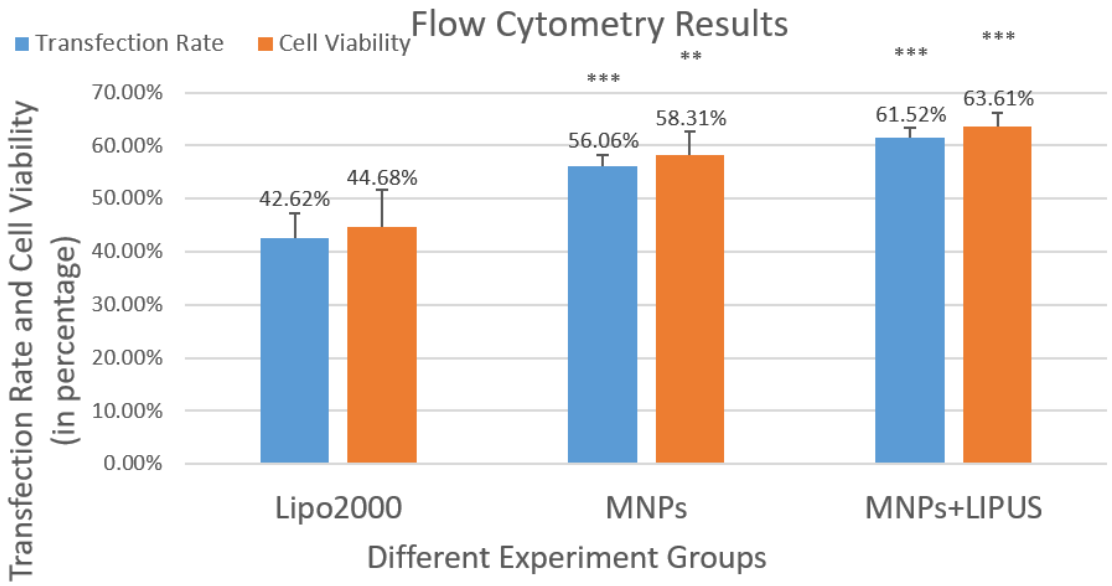
Compared to the standard Lipofectamine 2000 reagent mentioned above, the MNPs alone and MNPs coupled with LIPUS stimulation gave us better results (shown in Fig 4.5 a). They increased the transfection efficiency 1.3- and 1.45- fold, respectively, over the Lipofectamine 2000. Fig 4.5 c is the histogram plot of flow cytometry, showing the GFP fluorescence of the experimental group transfected with GFP plasmid, MNPs, and external magnetic field. The transfection efficiency was 57.503%, 13% higher than the Lipofectamine 2000, which indicated that the MNPs themselves could perform

better than the Lipofectamine under the external magnetic field. Fig 4.5 d is the histogram plot of cytometry by combining the MNPs and the LIPUS stimulation with the transfection rate of 61.52%. By synergistically combining MNPs and LIPUS, we can achieve efficiently and targeted gene delivery. We also ran an experiment with MNPs, but without the application of the external magnetic field and the transfection efficiency was only 14%, as shown in Fig 4.5 e. This is consistent with the idea that MNPs, as gene carriers, cannot efficiently deliver the material without the external magnet field targeting.

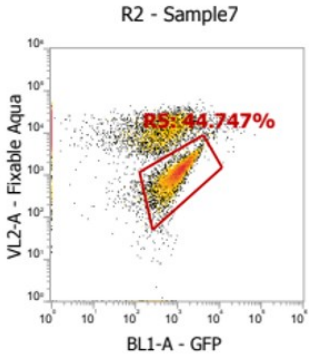
For the development of the additional LIPUS stimulation step of our method, we first selected the ultrasound condition using 10 minutes of treatment with an intensity of 30 mW/cm^2. We then performed the transfection using our experimental setup with the LIPUS device. We got an average transfection result of 61.52%, showing the highest transfection efficiency among our samples with the p-value of 0.0001. As a control, cells mixed with the genetic material (plasmid of interests) and treated with ultrasound were used to evaluate the effect of just LIPUS stimulation on the transfection without the MNPs/magnetic field. In that experiment, the transfection rate was only at 1%, as shown in Fig 4.5 f, indicating that LIPUS alone could not transfect the cells without the carriers. Our results showed that, though LIPUS waves could not function as a tool for transfection itself, they could permeabilize cell membranes and, coupled with another tool (i.e., MNPs in this case), could enhance gene delivery into the cells.

4.3.4 Cell Toxicity Results

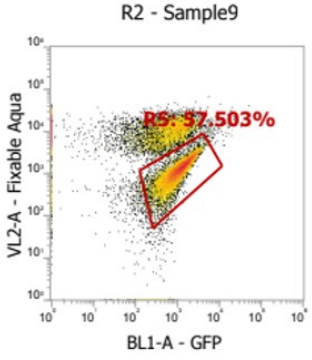
In our experiment, we used viability assay (Zombie Aqua) to evaluate the viability of HEK cells in our tests with different gene delivery approaches. Fig 4.5 g, 4.5 h, 4.5 i show three different groups of cell viability results (Group 1: Lipofectamine 2000, Group 2: MNPs/magnetic field without LIPUS stimulation, Group 3: MNPs/magnetic field plus LIPUS), and the overall results are showing in Fig 4.5 a. The negative control group (just untreated cells) showed viability at 91.524%, which was used as the background for all the results. Lipofectamine 2000 gave us 44.68% cell viability. MNPs showed 14% higher cell viability than the Lipofectamine 2000, indicating that our MNPs had lower cytotoxicity. The addition of the LIPUS device stimulation to the MNPs delivery further increased cell viability by up to 63.61% after the transfection with the p-value of 0.0002. These results showed that the LIPUS wave could stimulate cell growth and enhance cell viability.


Flow Cytometry Results
Transfection Rate
Cell Viability
Transfection Rate and Cell Viability (in percentage)
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
42.62%
44.68%
56.06%
58.31%
61.52%
63.61%
***
**
***
***
Lipo2000
MNPs
MNPs+LIPUS
Different Experiment Groups

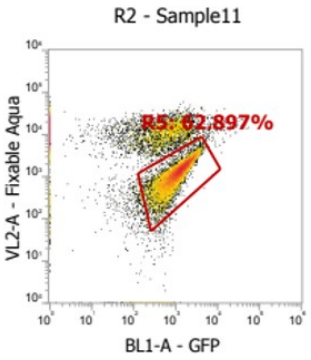
(a)


R2 - Sample7
R5: 44.747%
VL2-A - Fixable Aqua
BL1-A - GFP

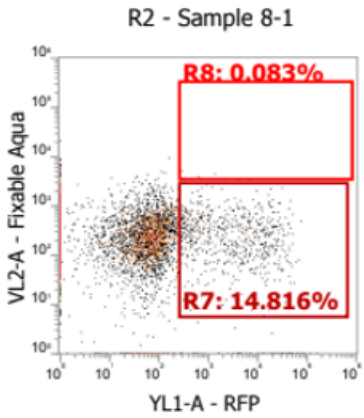
(b)


R2 - Sample9
R5: 57.503%
VL2-A - Fixable Aqua
BL1-A - GFP

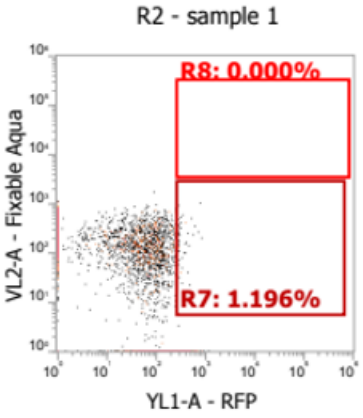
(c)


R2 - Sample11
R5: 62.897%
VL2-A - Fixable Aqua
BL1-A - GFP

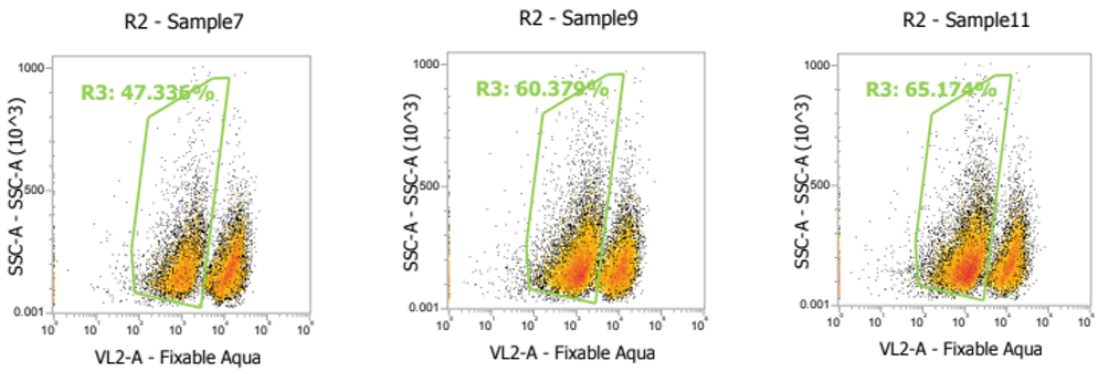
(d)


R2 - Sample 8-1
R8: 0.083%
R7: 14.816%
VL2-A - Fixable Aqua
YL1-A - RFP

(e)


R2 - sample 1
R8: 0.000%
R7: 1.196%
VL2-A - Fixable Aqua
YL1-A - RFP

(f)



(g) (h) (i)

Figure 4. 5 Quantification of transfection: (a) Overall transfection efficiency and cell viability results. (**:p<0.01, ***: p<0.001). Subfigures (b)-(i) show the flow cytometry histogram plots of transfection rates using GFP with different methods. (b) lipofectamine 2000, (c) our MNPs and magnet, p<0.001 (d) our suggested method: MNPs, magnet, in combination with LIPUS treatment, p<0.001. (e) MNPs only, (f) treated only with LIPUS. Cell viability results in the presence of Zombie Aqua viability dye when transfected with (g) lipofectamine 2000. (h) MNPs and magnet, p<0.01, (i) MNPs, magnet, and LIPUS, p<0.001.

4.3.5 Confocal Microscope Results

For confocal microscopy, the cells were transfected using the same protocol, but they were cultured on slides instead of a 12-well plate. Cells were stained with DAPI to evaluate the location of the gene in the transfected cells. Fig 4.6 shows the confocal microscopy results of the cells after transfection within 48 hours. In the figure, both Cherry Red plasmid and GFP plasmid HEK cells were sufficiently transfected using our developed technique. The merged images indicate genes mostly accumulated and

expressed in the nucleus, where the Dapi stains, confirming the successful delivery of the genetic material to the nucleus.

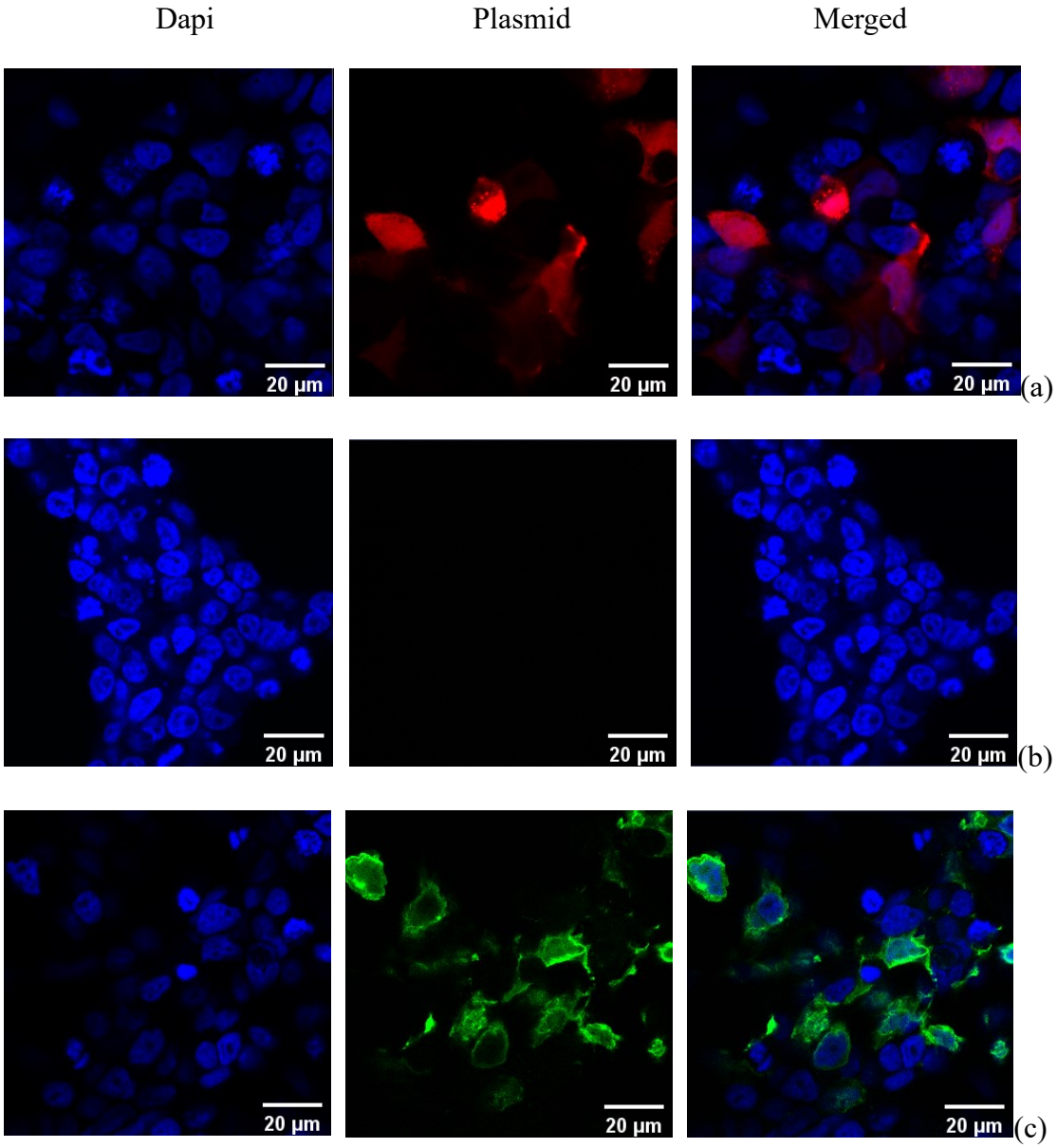


Figure 4. 6 Fluorescent images of cells transfected different plasmids using both MNPs and LIPUS stained with DAPI, (a) The control group (just cells), (b) GFP. (c) Cherry-red. Scale bars=20 µm.

4.4 Conclusions

In this study, we combined the application of an external magnetic field with MNPs and LIPUS stimulation for gene delivery. The uniqueness of our design is that, in addition to MNPs, used as a transfection carrier, we used the LIPUS cell stimulation to enhance gene delivery through increased cell permeability. In our experiments, we did the transfection on the HEK cells using our nanoparticles and got a 14% higher transfection rate, compared to the Lipofectamine 2000. The transfection efficiency further increases by 5%, when we add the LIPUS cell stimulation to the whole system, which was in line with our expectations. As for cell viability, Lipofectamine is known for its cytotoxicity and showed only 44.48% cell viability in our transfection experiments. A higher percentage of cells were alive after transfection when we used the MNPs with viability up to 58.31%. LIPUS stimulation added as an extra step during the MNPs transfection yielded even higher cell viability at 63.61%, compared to the MNPs only.

It is worth mentioning that our results and the Lipofectamine 2000 results were compared in terms of both the transfection efficiency and cell viability, and our technique showed better performance. LIPUS was shown to promote cell permeability and let the MNPs-DNA complex pass through and thus increasing the transfection efficiency and enhancing the cell viability. Because our assay is 10x cheaper than Lipofectamine 2000 and is also a chemical-based physical delivery approach, it can be

an attractive gene-delivery method for other hard-to-transfect cells (such as primary cells and neuron cells) and *in vivo*.

5 Conclusions and Future Work

5.1 Conclusions

In this thesis, two different automated disease diagnosis methods and one LIPUS based magnetic nanoparticle gene delivery system were proposed, which opened up a promising interdisciplinary research field in biomedical engineering applications. Chapter 2 designed the multimodal fusion model comprised of text, audio, and video for both depression detection and assessment tasks. Experiments on the DAIC-WOZ dataset showed a great improvement in performance, with a weighted F1 score of 0.85, an RMSE of 5.57, and an MAE of 4.48. The proposed model outperforms the baseline in both depression detection and assessment tasks and has comparable performance with other existing state-of-the-art depression detection methods. The empirical results show that compared with the unimodal model, the use of the multimodal model provides a better representation for depression, thereby improving the automated depression detection and assessment system. This research will help develop an automated depression detection system that combines various modalities and can be easily transferred to high-performance, portable, low-cost, and rapid depression diagnosis and prognosis devices.

Chapter 3 used high-resolution LC-MS to screen 191 blood samples and discovered Kyn, Trp and their ratio, IDO are excellent TB biomarkers. There was a significant difference in the concentration of these metabolites among different groups.

We employed the logistic regression algorithm to detect pulmonary TB and the AUC score, accuracy can be as high as 1.00 and 96% for classifying HC vs ATB and LTBI vs ATB. When we used IDO and t-spot to distinguish between NTB and ATB, the accuracy can always be above 80% both on the validation set and external independent cohort. The AUC performance of multi-class classification is generally greater than 0.83 and is 0.97 for the ATB class, especially. We conducted this study to propose and test Kyn, Trp, and IDO activity as novel biomarker indicators for the detection of ATB with the help of the t-spot. This study is only a pilot study, and non-targeted metabolomic is needed to be performed to add more significant biomarkers to enhance the multi-class classification. This study can contribute to developing diagnostic procedures in combination with other biomarkers and can be easily transferred into a high-performance, low-cost, noninvasive, and rapid pulmonary TB diagnosis and prognosis device.

Chapter 4 proposed a novel gene delivery system that synergistically combines non-viral chemical materials, MNPs, and physical technique, LIPUS, to achieve efficiently and targeted gene delivery. The uniqueness of our design is that, in addition to MNPs, used as a transfection carrier, we used the LIPUS cell stimulation to enhance gene delivery through increased cell permeability. In our experiments, we did the transfection on the HEK cells using our nanoparticles and got a 14% higher transfection rate, compared to the Lipofectamine 2000. The transfection efficiency further increases by 5%, when we add the LIPUS cell stimulation to the whole system, which was in line

with our expectations. As for cell viability, Lipofectamine is known for its cytotoxicity and showed only 44.48% cell viability in our transfection experiments. A higher percentage of cells were alive after transfection when we used the MNPs with viability up to 58.31%. LIPUS stimulation added as an extra step during the MNPs transfection yielded even higher cell viability at 63.61%, compared to the MNPs only. This new gene-delivery system is affordable, targeted, low-toxicity, yet high transfection efficiency, compared to other conventional approaches.

5.2 Future Work

Several directions of future work for these biomedical engineering applications should be mentioned. First of all, for the automated depression detection and assessment system, we currently used the video features provided by the dataset, instead of exploring more significant features from the raw video. However, video features play a key role in modeling the deep correlation between depression and facial emotions and patients with depression often display distorted facial expressions. So, if the raw video can be included in the dataset, the performance can be improved for sure. In addition, we need to collect an independent cohort to further validate our results to see whether our model can have a good generalization and perform well on unseen patients. For the automated TB diagnosis, for now, we just used three biomarkers, Kyn, Trp, and IDO. We believe that if more biomarkers are included, the performance of the multi-class classification and the classifications among HC, LTBI, and NTB can also be satisfied.

For the high transfection rate, low cytotoxicity gene delivery system, we now employed the gene delivery on the HEK cell, which is a kind of easy-transfected cell. The future work is to test and improve our gene delivery system on other hard-transfected cells (such as prime cells and neuron cells) and *in vivo*.

Reference

[1] J. McCarthy, “What is artificial intelligence?,” 2007.

[2] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, “Causability and explainability of artificial intelligence in medicine,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 4, p. e1312, 2019.

[3] D. Poole, A. Mackworth, and R. Goebel, “Computational Intelligence,” 1998.

[4] S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach. Third Edit,” *Prentice Hall. doi*, vol. 10, pp. B978-012161964, 2010.

[5] Y. Mintz and R. Brodie, “Introduction to artificial intelligence in medicine,” *Minimally Invasive Therapy & Allied Technologies*, vol. 28, no. 2, pp. 73–81, 2019.

[6] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, “How to grow a mind: Statistics, structure, and abstraction,” *science*, vol. 331, no. 6022, pp. 1279–1285, 2011.

[7] A. Newell, J. C. Shaw, and H. A. Simon, “Report on a general problem solving program,” in *IFIP congress*, 1959, vol. 256, p. 64.

[8] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[9] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[10] X.-D. Zhang, “Machine learning,” in *A Matrix Algebra Approach to Artificial Intelligence*, Springer, 2020, pp. 223–440.

[11] P. Ongsulee, “Artificial intelligence, machine learning and deep learning,” in *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, 2017, pp. 1–6.

[12] L. Gomes, “Machine-learning maestro michael jordan on the delusions of big data and other huge engineering efforts,” *IEEE spectrum*, vol. 20, 2014.

[13] R. Collobert, “Deep learning for efficient discriminative parsing,” in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 2011, pp. 224–232.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[15] L. B. Lusted, “Medical electronics,” *New England Journal of Medicine*, vol. 252, no. 14, pp. 580–585, 1955.

[16] R. S. Ledley and L. B. Lusted, “Reasoning foundations of medical diagnosis,” *Science*, vol. 130, no. 3366, pp. 9–21, 1959.

[17] A. A. Gunn, “The diagnosis of acute abdominal pain with computer analysis.,” *Journal of the Royal College of Surgeons of Edinburgh*, vol. 21, no.

3, pp. 170–172, 1976.

[18] A. N. Ramesh, C. Kambhampati, J. R. T. Monson, and P. J. Drew, “Artificial intelligence in medicine.,” *Annals of the Royal College of Surgeons of England*, vol. 86, no. 5, p. 334, 2004.

[19] I. Kononenko, “Machine learning for medical diagnosis: history, state of the art and perspective,” *Artificial Intelligence in medicine*, vol. 23, no. 1, pp. 89–109, 2001.

[20] L. Strohm, C. Hehakaya, E. R. Ranschaert, W. P. C. Boon, and E. H. M. Moors, “Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors,” *European radiology*, vol. 30, pp. 5525–5532, 2020.

[21] J. H. Thrall *et al.*, “Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success,” *Journal of the American College of Radiology*, vol. 15, no. 3, pp. 504–508, 2018.

[22] D. A. Bluemke, “Radiology in 2018: are you working with AI or being replaced by AI?,” *Radiology*, vol. 287, no. 2, pp. 365–366, 2018.

[23] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[24] P. Teare, M. Fishman, O. Benzaquen, E. Toledano, and E. Elnekave, “Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement,” *Journal of*

digital imaging, vol. 30, no. 4, pp. 499–505, 2017.

[25] A. Bar, L. Wolf, O. B. Amitai, E. Toledano, and E. Elnekave, "Compression fractures detection on CT," in *Medical Imaging 2017: Computer-Aided Diagnosis*, 2017, vol. 10134, p. 1013440.

[26] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[27] K. R. Laukamp *et al.*, "Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI," *European radiology*, vol. 29, no. 1, pp. 124–132, 2019.

[28] B. E. Bejnordi *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[29] R. Li *et al.*, "Deep learning based imaging data completion for improved brain disease diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2014, pp. 305–312.

[30] Y.-H. Li, L. Zhang, Q.-M. Hu, H.-W. Li, F.-C. Jia, and J.-H. Wu, "Automatic subarachnoid space segmentation and hemorrhage detection in clinical head CT scans," *International journal of computer assisted radiology and surgery*, vol. 7, no. 4, pp. 507–516, 2012.

[31] K. Zhang *et al.*, "Clinically applicable AI system for accurate diagnosis,

quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography," *Cell*, vol. 181, no. 6, pp. 1423–1433, 2020.

[32] A. S. Becker, M. Marcon, S. Ghafoor, M. C. Wurnig, T. Frauenfelder, and A. Boss, "Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer," *Investigative radiology*, vol. 52, no. 7, pp. 434–440, 2017.

[33] H. Chougrad, H. Zouaki, and O. Alheyane, "Deep convolutional neural networks for breast cancer screening," *Computer methods and programs in biomedicine*, vol. 157, pp. 19–30, 2018.

[34] M. Ogino, Z. Li, and A. Shimizu, "Augmented Radiology: Feature Space Transfer Model for Prostate Cancer Stage Prediction," *IEEE Access*, vol. 9, pp. 102559–102566, 2021.

[35] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology," *Nature reviews Clinical oncology*, vol. 16, no. 11, pp. 703–715, 2019.

[36] Z. Dlamini, F. Z. Francies, R. Hull, and R. Marima, "Artificial intelligence (AI) and big data in cancer and precision oncology," *Computational and Structural Biotechnology Journal*, 2020.

[37] V. Y. Londhe and B. Bhasin, "Artificial intelligence and its potential in oncology," *Drug discovery today*, vol. 24, no. 1, pp. 228–232, 2019.

[38] A. Šećkanović *et al.*, “Review of artificial intelligence application in cardiology,” in *2020 9th Mediterranean Conference on Embedded Computing (MECO)*, 2020, pp. 1–5.

[39] A. Haleem, M. Javaid, R. P. Singh, and R. Suman, “Applications of Artificial Intelligence (AI) for cardiology during COVID-19 pandemic,” *Sustainable Operations and Computers*, vol. 2, pp. 71–78, 2021.

[40] F. Lopez-Jimenez *et al.*, “Artificial intelligence in cardiology: present and future,” in *Mayo Clinic Proceedings*, 2020, vol. 95, no. 5, pp. 1015–1039.

[41] C. le Berre *et al.*, “Application of artificial intelligence to gastroenterology and hepatology,” *Gastroenterology*, vol. 158, no. 1, pp. 76–94, 2020.

[42] Y. J. Yang and C. S. Bang, “Application of artificial intelligence in gastroenterology,” *World journal of gastroenterology*, vol. 25, no. 14, p. 1666, 2019.

[43] W. Lu, Y. Tong, Y. Yu, Y. Xing, C. Chen, and Y. Shen, “Applications of artificial intelligence in ophthalmology: general overview,” *Journal of ophthalmology*, vol. 2018, 2018.

[44] D. S. J. Ting *et al.*, “Artificial intelligence for anterior segment diseases: emerging applications in ophthalmology,” *British Journal of Ophthalmology*, vol. 105, no. 2, pp. 158–168, 2021.

[45] Z. Tan, J. Scheetz, and M. He, “Artificial intelligence in ophthalmology: accuracy, challenges, and clinical application,” *The Asia-Pacific Journal of*

Ophthalmology, vol. 8, no. 3, pp. 197–199, 2019.

[46] S. M. McKinney *et al.*, "International evaluation of an AI system for breast cancer screening," *Nature*, vol. 577, no. 7788, pp. 89–94, 2020.

[47] K.-S. Wang *et al.*, "Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence," *BMC medicine*, vol. 19, no. 1, pp. 1–12, 2021.

[48] Z. I. Attia *et al.*, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *The Lancet*, vol. 394, no. 10201, pp. 861–867, 2019.

[49] C.-C. Wu *et al.*, "An artificial intelligence approach to early predict non-ST-elevation myocardial infarction patients with chest pain," *Computer methods and programs in biomedicine*, vol. 173, pp. 109–117, 2019.

[50] Z. Tan, W. Wang, R. Zong, J. Pan, and H. Yang, "Classification of heart sound signals in congenital heart disease based on convolutional neural network," *Sheng wu yi xue gong cheng xue za zhi= Journal of biomedical engineering= Shengwu yixue gongchengxue zazhi*, vol. 36, no. 5, pp. 728–736, 2019.

[51] J. C. Peng, Z. H. Ran, and J. Shen, "Seasonal variation in onset and relapse of IBD and a model to predict the frequency of onset, relapse, and severity of IBD based on artificial neural network," *International journal of colorectal disease*, vol. 30, no. 9, pp. 1267–1273, 2015.

[52] F. Hardalaç, M. Basaranoglu, M. Yuksel, U. Kutbay, M. Kaplan, and Y. Ozderin Ozin, "The rate of mucosal healing by azathioprine therapy and prediction by artificial systems," *Turk. J. Gastroenterol*, vol. 26, no. 4, pp. 315–321, 2015.

[53] T. Takayama *et al.*, "Computer-aided prediction of long-term prognosis of patients with ulcerative colitis after cytoapheresis therapy," *PloS one*, vol. 10, no. 6, p. e0131197, 2015.

[54] G. Rotondano *et al.*, "Artificial neural networks accurately predict mortality in patients with nonvariceal upper GI bleeding," *Gastrointestinal Endoscopy*, vol. 73, no. 2, pp. 218–226, 2011.

[55] A. Das *et al.*, "Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model," *The Lancet*, vol. 362, no. 9392, pp. 1261–1266, 2003.

[56] E. Lahner *et al.*, "Possible contribution of artificial neural networks and linear discriminant analysis in recognition of patients with suspected atrophic body gastritis.," *World journal of gastroenterology*, vol. 11, no. 37, pp. 5867–5873, 2005.

[57] F. Pace *et al.*, "Artificial neural networks are able to recognize gastro-oesophageal reflux disease patients solely on the basis of clinical data," *European journal of gastroenterology & hepatology*, vol. 17, no. 6, pp. 605–610,

2005.

[58] T. K. Yoo and E.-C. Park, “Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study,” *BMC medical informatics and decision making*, vol. 13, no. 1, pp. 1–14, 2013.

[59] M. L. Ribeiro, S. G. Nunes, and J. G. Cunha-Vaz, “Microaneurysm turnover at the macula predicts risk of development of clinically significant macular edema in persons with mild nonproliferative diabetic retinopathy,” *Diabetes care*, vol. 36, no. 5, pp. 1254–1259, 2013.

[60] C. Raja and N. Gangatharan, “A hybrid swarm algorithm for optimizing glaucoma diagnosis,” *Computers in biology and medicine*, vol. 63, pp. 196–207, 2015.

[61] M. S. Haleem *et al.*, “Regional image features model for automatic classification between normal and glaucoma in fundus and scanning laser ophthalmoscopy (SLO) images,” *Journal of medical systems*, vol. 40, no. 6, p. 132, 2016.

[62] M. J. J. P. van Grinsven *et al.*, “Automatic identification of reticular pseudodrusen using multimodal retinal image analysis,” *Investigative ophthalmology & visual science*, vol. 56, no. 1, pp. 633–639, 2015.

[63] A. K. Feeny, M. Tadarati, D. E. Freund, N. M. Bressler, and P. Burlina, “Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images,” *Computers in biology and*

medicine, vol. 65, pp. 124–136, 2015.

[64] Q. Chen *et al.*, "Automated drusen segmentation and quantification in SD-OCT images," *Medical image analysis*, vol. 17, no. 8, pp. 1058–1072, 2013.

[65] X. Liu *et al.*, "Localization and diagnosis framework for pediatric cataracts based on slit-lamp images using deep features of a convolutional neural network," *PloS one*, vol. 12, no. 3, p. e0168606, 2017.

[66] X. Gao, S. Lin, and T. Y. Wong, "Automatic feature learning to grade nuclear cataracts based on deep learning," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 11, pp. 2693–2701, 2015.

[67] E. Long *et al.*, "An artificial intelligence platform for the multihospital collaborative management of congenital cataracts," *Nature biomedical engineering*, vol. 1, no. 2, pp. 1–8, 2017.

[68] S. O'Sullivan *et al.*, "Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 15, no. 1, p. e1968, 2019.

[69] D. A. Hashimoto, G. Rosman, D. Rus, and O. R. Meireles, "Artificial intelligence in surgery: promises and perils," *Annals of surgery*, vol. 268, no. 1, p. 70, 2018.

[70] E. M. Bonrath, L. E. Gordon, and T. P. Grantcharov, "Characterising 'near miss' events in complex laparoscopic surgery through video analysis," *BMJ*

quality & safety, vol. 24, no. 8, pp. 516–521, 2015.

[71] M. Volkov, D. A. Hashimoto, G. Rosman, O. R. Meireles, and D. Rus, "Machine learning and coresets for automated real-time video segmentation of laparoscopic and robot-assisted surgery," in *2017 IEEE international conference on robotics and automation (ICRA)*, 2017, pp. 754–759.

[72] P. Natarajan, J. C. Frenzel, and D. H. Smaltz, *Demystifying big data and machine learning for healthcare*. CRC Press, 2017.

[73] T. R. Grenda, J. C. Pradarelli, and J. B. Dimick, "Using surgical video to improve technique and skill," *Annals of surgery*, vol. 264, no. 1, p. 32, 2016.

[74] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017.

[75] G. Shen *et al.*, "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution.," in *IJCAI*, 2017, pp. 3838–3844.

[76] S. Smys and J. S. Raj, "Analysis of Deep Learning Techniques for Early Detection of Depression on Social Media Network-A Comparative Study," *Journal of trends in Computer Science and Smart technology (TCSST)*, vol. 3, no. 01, pp. 24–39, 2021.

[77] M. Valstar *et al.*, "Avec 2013: the continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013, pp. 3–10.

[78] M. Valstar *et al.*, “Avec 2016: Depression, mood, and emotion recognition workshop and challenge,” in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 3–10.

[79] F. Ringeval *et al.*, “Avec 2017: Real-life depression, and affect recognition workshop and challenge,” in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, 2017, pp. 3–9.

[80] F. Ringeval *et al.*, “AVEC 2019 workshop and challenge: state-of-mind, detecting depression with AI, and cross-cultural affect recognition,” in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 3–12.

[81] M. Rohanian, J. Hough, and M. Purver, “Detecting Depression with Word-Level Multimodal Fusion.,” in *INTERSPEECH*, 2019, pp. 1443–1447.

[82] M. Niu, J. Tao, B. Liu, J. Huang, and Z. Lian, “Multimodal spatiotemporal representation for automatic depression level detection,” *IEEE Transactions on Affective Computing*, 2020.

[83] Y. Gong and C. Poellabauer, “Topic modeling based multi-modal depression detection,” in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*, 2017, pp. 69–76.

[84] M. Z. Uddin, K. K. Dysthe, A. Følstad, and P. B. Brandtzaeg, “Deep learning for prediction of depressive symptoms in a large textual dataset,” *Neural Computing and Applications*, pp. 1–24, 2021.

[85] S. Jaeger *et al.*, "Automatic screening for tuberculosis in chest radiographs: a survey," *Quantitative imaging in medicine and surgery*, vol. 3, no. 2, p. 89, 2013.

[86] T. Xu, I. Cheng, R. Long, and M. Mandal, "Novel coarse-to-fine dual scale technique for tuberculosis cavity detection in chest radiographs," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 1–18, 2013.

[87] Y. Yang and Y.-L. Song, "Localization algorithm and implementation for focal of pulmonary tuberculosis chest image," in *2010 International Conference on Machine Vision and Human-machine Interface*, 2010, pp. 361–364.

[88] J. C. Schoeman and I. du Preez, "A comparison of four sputum pre-extraction preparation methods for identifying and characterising Mycobacterium tuberculosis using GCxGC-TOFMS metabolomics," *Journal of microbiological methods*, vol. 91, no. 2, pp. 301–311, 2012.

[89] I. du Preez and D. T. Loots, "New sputum metabolite markers implicating adaptations of the host to Mycobacterium tuberculosis, and vice versa," *Tuberculosis*, vol. 93, no. 3, pp. 330–337, 2013.

[90] S. Feng, Y.-Q. Du, L. Zhang, L. Zhang, R.-R. Feng, and S.-Y. Liu, "Analysis of serum metabolic profile by ultra-performance liquid chromatography-mass spectrometry for biomarkers discovery: application in a pilot study to discriminate patients with tuberculosis," *Chinese medical journal*, vol. 128, no. 2, p. 159, 2015.

[91] N. Che *et al.*, “Decreased serum 5-oxoproline in TB patients is associated with pathological damage of the lung,” *Clinica Chimica Acta*, vol. 423, pp. 5–9, 2013.

[92] M. Phillips *et al.*, “Breath biomarkers of active pulmonary tuberculosis,” *Tuberculosis*, vol. 90, no. 2, pp. 145–151, 2010.

[93] A. H. J. Kolk *et al.*, “Breath analysis as a potential diagnostic tool for tuberculosis,” *The International Journal of Tuberculosis and Lung Disease*, vol. 16, no. 6, pp. 777–782, 2012.

[94] S. Mahapatra *et al.*, “A metabolic biosignature of early response to anti-tuberculosis treatment,” *BMC infectious diseases*, vol. 14, no. 1, pp. 1–11, 2014.

[95] R. C. Mulligan, “The basic science of gene therapy,” *Science*, vol. 260, no. 5110, pp. 926–932, 1993.

[96] S.-D. Li and L. Huang, “Gene therapy progress and prospects: non-viral gene therapy by systemic delivery,” *Gene therapy*, vol. 13, no. 18, pp. 1313–1319, 2006.

[97] C. E. Walsh, “Gene therapy progress and prospects: gene therapy for the hemophilias,” *Gene therapy*, vol. 10, no. 12, pp. 999–1003, 2003.

[98] J. C. T. van Deutekom and G.-J. B. van Ommen, “Advances in Duchenne muscular dystrophy gene therapy,” *Nature Reviews Genetics*, vol. 4, no. 10, pp. 774–783, 2003.

[99] S. Ferrari, D. M. Geddes, and E. W. F. W. Alton, “Barriers to and new

approaches for gene therapy and gene delivery in cystic fibrosis," *Advanced drug delivery reviews*, vol. 54, no. 11, pp. 1373–1393, 2002.

[100] V. J. Dzau, K. Beatt, G. Pompilio, and K. Smith, "Current perceptions of cardiovascular gene therapy," *The American journal of cardiology*, vol. 92, no. 9, pp. 18–23, 2003.

[101] E. A. Burton, J. C. Glorioso, and D. J. Fink, "Gene therapy progress and prospects: Parkinson's disease," *Gene Therapy*, vol. 10, no. 20, pp. 1721–1727, 2003.

[102] J. M. Alisky and B. L. Davidson, "Gene therapy for amyotrophic lateral sclerosis and other motor neuron diseases," *Human gene therapy*, vol. 11, no. 17, pp. 2315–2329, 2000.

[103] M. H. Tuszynski, "Growth-factor gene therapy for neurodegenerative disorders," *The Lancet Neurology*, vol. 1, no. 1, pp. 51–57, 2002.

[104] B. A. Bunnell and R. A. Morgan, "Gene therapy for infectious diseases," *Clinical microbiology reviews*, vol. 11, no. 1, pp. 42–56, 1998.

[105] R. G. Vile, S. J. Russell, and N. R. Lemoine, "Cancer gene therapy: hard lessons and new courses," *Gene therapy*, vol. 7, no. 1, pp. 2–8, 2000.

[106] D. Kerr, "Clinical development of gene therapy for colorectal cancer," *Nature Reviews Cancer*, vol. 3, no. 8, pp. 615–622, 2003.

[107] I. A. McNeish, S. J. Bell, and N. R. Lemoine, "Gene therapy progress and prospects: cancer gene therapy using tumour suppressor genes," *Gene therapy*,

vol. 11, no. 6, pp. 497–503, 2004.

[108] D. W. Pack, A. S. Hoffman, S. Pun, and P. S. Stayton, “Design and development of polymers for gene delivery,” *Nature reviews Drug discovery*, vol. 4, no. 7, pp. 581–593, 2005.

[109] I. M. Verma *et al.*, “Gene therapy: promises, problems and prospects,” in *Genes and Resistance to Disease*, Springer, 2000, pp. 147–157.

[110] Y. K. Sung and S. W. Kim, “Recent advances in the development of gene delivery systems,” *Biomaterials research*, vol. 23, no. 1, pp. 1–7, 2019.

[111] C. H. Evans, “Gene therapy for bone healing,” *Expert reviews in molecular medicine*, vol. 12, 2010.

[112] J. M. Davidson, “First-class delivery: getting growth factors to their destination,” *Journal of Investigative Dermatology*, vol. 128, no. 6, pp. 1360–1362, 2008.

[113] S. C. Semple, T. O. Harasym, K. A. Clow, S. M. Ansell, S. K. Klimuk, and M. J. Hope, “Immunogenicity and rapid blood clearance of liposomes containing polyethylene glycol-lipid conjugates and nucleic acid,” *Journal of Pharmacology and Experimental Therapeutics*, vol. 312, no. 3, pp. 1020–1026, 2005.

[114] L. Naldini, “Gene therapy returns to centre stage,” *Nature*, vol. 526, no. 7573, pp. 351–360, 2015.

[115] M. Elsabahy, A. Nazarali, and M. Foldvari, “Non-viral nucleic acid

delivery: key challenges and future directions," *Current drug delivery*, vol. 8, no. 3, pp. 235–244, 2011.

[116] D. Luo and W. M. Saltzman, "Synthetic DNA delivery systems," *Nature biotechnology*, vol. 18, no. 1, pp. 33–37, 2000.

[117] I. Roy, S. Mitra, A. Maitra, and S. Mozumdar, "Calcium phosphate nanoparticles as novel non-viral vectors for targeted gene delivery," *International journal of pharmaceutics*, vol. 250, no. 1, pp. 25–33, 2003.

[118] M. E. Davis, "Non-viral gene delivery systems," *Current opinion in biotechnology*, vol. 13, no. 2, pp. 128–131, 2002.

[119] P. Belguise-Valladier and J.-P. Behr, "Nonviral gene delivery: towards artificial viruses," *Cytotechnology*, vol. 35, no. 3, pp. 197–201, 2001.

[120] D. K. KHOSRAVI, M. R. Mozafari, L. Rashidi, and M. Mohammadi, "Calcium based non-viral gene delivery: an overview of methodology and applications," 2010.

[121] M. Ravi Kumar, G. Hellermann, R. F. Lockey, and S. S. Mohapatra, "Nanoparticle-mediated gene delivery: state of the art," *Expert opinion on biological therapy*, vol. 4, no. 8, pp. 1213–1224, 2004.

[122] E. Neumann, "Electric gene transfer into culture cells," *Bioelectrochemistry and Bioenergetics*, vol. 13, no. 1–3, pp. 219–223, 1984.

[123] Y. Liu, J. Yan, and M. R. Prausnitz, "Can ultrasound enable efficient intracellular uptake of molecules? A retrospective literature review and analysis,"

Ultrasound in medicine & biology, vol. 38, no. 5, pp. 876–888, 2012.

[124] J. A. Wolff *et al.*, "Direct gene transfer into mouse muscle in vivo," *Science*, vol. 247, no. 4949, pp. 1465–1468, 1990.

[125] M. Fechheimer, J. F. Boylan, S. Parker, J. E. Sisken, G. L. Patel, and S. G. Zimmer, "Transfection of mammalian cells with plasmid DNA by scrape loading and sonication loading," *Proceedings of the National Academy of Sciences*, vol. 84, no. 23, pp. 8463–8467, 1987.

[126] H. J. Kim, J. F. Greenleaf, R. R. Kinnick, J. T. Bronk, and M. E. Bolander, "Ultrasound-mediated transfection of mammalian cells," *Human gene therapy*, vol. 7, no. 11, pp. 1339–1346, 1996.

[127] S. Bao, B. D. Thrall, and D. L. Miller, "Transfection of a reporter plasmid into cultured cells by sonoporation in vitro," *Ultrasound in medicine & biology*, vol. 23, no. 6, pp. 953–959, 1997.

[128] A. Lawrie *et al.*, "Ultrasound enhances reporter gene expression after transfection of vascular cells in vitro," *Circulation*, vol. 99, no. 20, pp. 2617–2620, 1999.

[129] D. B. Tata, F. Dunn, and D. J. Tindall, "Selective clinical ultrasound signals mediate differential gene transfer and expression in two human prostate cancer cell lines: LnCap and PC-3," *Biochemical and biophysical research communications*, vol. 234, no. 1, pp. 64–67, 1997.

[130] D. Ensminger and L. J. Bond, *Ultrasonics: fundamentals, technologies,*

and applications. CRC press, 2011.

[131] R. W. Wood and A. L. Loomis, "XXXVIII. The physical and biological effects of high-frequency sound-waves of great intensity," *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, vol. 4, no. 22, pp. 417–436, 1927.

[132] S. Kothari and V. Kaul, "Therapeutic applications of endoscopic ultrasound," *Ultrasound Clinics*, vol. 9, no. 1, pp. 53–65, 2014.

[133] G. Lin, A. B. Reed-Maldonado, M. Lin, Z. Xin, and T. F. Lue, "Effects and mechanisms of low-intensity pulsed ultrasound for chronic prostatitis and chronic pelvic pain syndrome," *International journal of molecular sciences*, vol. 17, no. 7, p. 1057, 2016.

[134] N. M. Pounder and A. J. Harrison, "Low intensity pulsed ultrasound for fracture healing: a review of the clinical evidence and the associated biological mechanism of action," *Ultrasonics*, vol. 48, no. 4, pp. 330–338, 2008.

[135] G. ter Haar, "Therapeutic ultrasound," *European Journal of ultrasound*, vol. 9, no. 1, pp. 3–9, 1999.

[136] K. N. Malizos, M. E. Hantes, V. Protopappas, and A. Papachristos, "Low-intensity pulsed ultrasound for bone healing: an overview," *Injury*, vol. 37, no. 1, pp. S56–S62, 2006.

[137] Z. Xin, G. Lin, H. Lei, T. F. Lue, and Y. Guo, "Clinical applications of low-intensity pulsed ultrasound and its potential role in urology," *Translational*

andrology and urology, vol. 5, no. 2, p. 255, 2016.

[138] A. Khanna, R. T. C. Nelmes, N. Gougoulias, N. Maffulli, and J. Gray, "The effects of LIPUS on soft-tissue healing: a review of literature," *British medical bulletin*, vol. 89, no. 1, pp. 169–182, 2009.

[139] X. Jiang *et al.*, "A review of low-intensity pulsed ultrasound for therapeutic applications," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 10, pp. 2704–2718, 2018.

[140] J. Xing *et al.*, "Increasing vaccine production using pulsed ultrasound waves," *PloS one*, vol. 12, no. 11, p. e0187048, 2017.

[141] O. Savchenko *et al.*, "Algal cell response to pulsed waved stimulation and its application to increase algal lipid production," *Scientific reports*, vol. 7, no. 1, pp. 1–13, 2017.

[142] P. Xu *et al.*, "Low-intensity pulsed ultrasound-mediated stimulation of hematopoietic stem/progenitor cell viability, proliferation and differentiation in vitro," *Biotechnology letters*, vol. 34, no. 10, pp. 1965–1973, 2012.

[143] J. Chen, X. James, W. T. Ang, and H. Gul, "Enhanced animal cell growth using ultrasound." Google Patents, Feb. 24, 2015.

[144] Y. Zhao, J. Xing, J. Z. Xing, W. T. Ang, and J. Chen, "Applications of low-intensity pulsed ultrasound to increase monoclonal antibody production in CHO cells using shake flasks or wavebags," *Ultrasonics*, vol. 54, no. 6, pp. 1439–1447, 2014.

[145] C. He *et al.*, “Microbubble-enhanced cell membrane permeability in high gravity field,” *Cellular and Molecular Bioengineering*, vol. 6, no. 3, pp. 266–278, 2013.

[146] K. G. Baker, V. J. Robertson, and F. A. Duck, “A review of therapeutic ultrasound: biophysical effects,” *Physical therapy*, vol. 81, no. 7, pp. 1351–1358, 2001.

[147] World Health Organization, “Depression.” Aug. 2019. [Online]. Available: https://www.who.int/health-topics/depression

[148] M. Nasir, A. Jati, P. G. Shivakumar, S. Nallan Chakravarthula, and P. Georgiou, “Multimodal and multiresolution depression detection from speech and facial landmark features,” in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 43–50.

[149] World Health Organization, “Depression and other common mental disorders: global health estimates,” World Health Organization, 2017.

[150] M. Fava and K. S. Kendler, “Major depressive disorder,” *Neuron*, vol. 28, no. 2, pp. 335–341, 2000.

[151] World Health Organization, “World health statistics 2021,” World Health Organization, 2021.

[152] J.-P. Lépine and M. Briley, “The increasing burden of depression,” *Neuropsychiatric disease and treatment*, vol. 7, no. Suppl 1, p. 3, 2011.

[153] K. Hawton, C. C. i Comabella, C. Haw, and K. Saunders, “Risk factors for

suicide in individuals with depression: a systematic review," *Journal of affective disorders*, vol. 147, no. 1–3, pp. 17–28, 2013.

[154] K. Kroenke, T. W. Strine, R. L. Spitzer, J. B. W. Williams, J. T. Berry, and A. H. Mokdad, "The PHQ-8 as a measure of current depression in the general population," *Journal of affective disorders*, vol. 114, no. 1–3, pp. 163–173, 2009.

[155] M. Carey *et al.*, "Accuracy of general practitioner unassisted detection of depression," *Australian & New Zealand Journal of Psychiatry*, vol. 48, no. 6, pp. 571–578, 2014.

[156] A. Ray, S. Kumar, R. Reddy, P. Mukherjee, and R. Garg, "Multi-level attention network using text, audio and video for depression prediction," in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 81–88.

[157] S. Scherer *et al.*, "Automatic audiovisual behavior descriptors for psychological disorder analysis," *Image and Vision Computing*, vol. 32, no. 10, pp. 648–658, 2014.

[158] J. F. Cohn *et al.*, "Detecting depression from facial actions and vocal prosody," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1–7.

[159] S. Rude, E.-M. Gortner, and J. Pennebaker, "Language use of depressed and depression-vulnerable college students," *Cognition & Emotion*, vol. 18, no. 8, pp. 1121–1133, 2004.

[160] P. Resnik, A. Garron, and R. Resnik, “Using topic modeling to improve prediction of neuroticism and depression in college students,” in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1348–1353.

[161] C. Y. Chiu, H. Y. Lane, J. L. Koh, and A. L. P. Chen, “Multimodal depression detection on instagram considering time interval of posts,” *Journal of Intelligent Information Systems*, vol. 56, no. 1, pp. 25–47, 2021.

[162] H. Ansari, A. Vijayvergia, and K. Kumar, “DCR-HMM: Depression detection based on Content Rating using Hidden Markov Model,” in *2018 Conference on Information and Communication Technology (CICT)*, 2018, pp. 1–6.

[163] L. Tong, Q. Zhang, A. Sadka, L. Li, and H. Zhou, “Inverse boosting pruning trees for depression detection on Twitter,” *arXiv preprint arXiv:1906.00398*, 2019.

[164] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, “Deep learning for depression detection of twitter users,” in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 2018, pp. 88–97.

[165] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, “Depression detection from social network data using machine learning techniques,” *Health information science and systems*, vol. 6, no. 1, pp. 1–12,

2018.

[166] S. Yin, C. Liang, H. Ding, and S. Wang, “A multi-modal hierarchical recurrent neural network for depression detection,” in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 65–71.

[167] M. Niu, K. Chen, Q. Chen, and L. Yang, “HCAG: A Hierarchical Context-Aware Graph Attention Model for Depression Detection,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 4235–4239.

[168] T. al Hanai, M. M. Ghassemi, and J. R. Glass, “Detecting Depression with Audio/Text Sequence Modeling of Interviews.,” in *Interspeech*, 2018, pp. 1716–1720.

[169] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, “A review of depression and suicide risk assessment using speech analysis,” *Speech Communication*, vol. 71, pp. 10–49, 2015.

[170] M. Rodrigues Makiuchi, T. Warnita, K. Uto, and K. Shinoda, “Multimodal fusion of bert-cnn and gated cnn representations for depression detection,” in *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*, 2019, pp. 55–63.

[171] S. Amiriparian *et al.*, “Snore sound classification using image-based deep spectrum features,” 2017.

[172] Y. Aytar, C. Vondrick, and A. Torralba, “Soundnet: Learning sound representations from unlabeled video,” *Advances in neural information processing systems*, vol. 29, pp. 892–900, 2016.

[173] Z. Huang, M. Dong, Q. Mao, and Y. Zhan, “Speech emotion recognition using CNN,” in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 801–804.

[174] X. Ma, H. Yang, Q. Chen, D. Huang, and Y. Wang, “Depaudionet: An efficient deep model for audio based depression classification,” in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 35–42.

[175] M. Muzammel, H. Salam, Y. Hoffmann, M. Chetouani, and A. Othmani, “AudVowelConsNet: A phoneme-level based deep CNN architecture for clinical depression diagnosis,” *Machine Learning with Applications*, vol. 2, p. 100005, 2020.

[176] A. Haque, M. Guo, A. S. Miner, and L. Fei-Fei, “Measuring depression symptom severity from spoken language and 3D facial expressions,” *arXiv preprint arXiv:1811.08592*, 2018.

[177] L. He and C. Cao, “Automated depression analysis using convolutional neural networks from speech,” *Journal of biomedical informatics*, vol. 83, pp. 103–111, 2018.

[178] G. Lam, H. Dongyan, and W. Lin, “Context-aware deep learning for multi-

modal depression detection," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 3946–3950.

[179] A. Pampouchidou *et al.*, "Depression assessment by fusing high and low level features from audio, video, and text," in *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*, 2016, pp. 27–34.

[180] A. Bobick and J. Davis, "Real-time recognition of activity using temporal templates," in *Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96*, 1996, pp. 39–42.

[181] H. Meng, D. Huang, H. Wang, H. Yang, M. Ai-Shuraifi, and Y. Wang, "Depression recognition based on dynamic facial and vocal expression features using partial least square regression," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, 2013, pp. 21–30.

[182] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–10.

[183] A. Pampouchidou *et al.*, "Facial geometry and speech analysis for depression detection," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 1433–1436.

[184] Y. Wang *et al.*, "Automatic Depression Detection via Facial Expressions

Using Multiple Instance Learning," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1933–1936.

[185] M. Kächele, M. Glodek, D. Zharkov, S. Meudt, and F. Schwenker, "Fusion of audio-visual features using hierarchical classifier systems for the recognition of affective states and the state of depression," *depression*, vol. 1, no. 1, pp. 671–678, 2014.

[186] L. Chao, J. Tao, M. Yang, and Y. Li, "Multi task sequence learning for depression scale prediction from video," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2015, pp. 526–531.

[187] J. Gratch *et al.*, "The distress analysis interview corpus of human and computer interviews.," in *LREC*, 2014, pp. 3123–3128.

[188] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*, 2014, pp. 1188–1196.

[189] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[190] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[191] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for

efficient text classification," *arXiv preprint arXiv:1607.01759*, 2016.

[192] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.

[193] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, "Supervised learning of universal sentence representations from natural language inference data," *arXiv preprint arXiv:1705.02364*, 2017.

[194] D. Cer *et al.*, "Universal sentence encoder," *arXiv preprint arXiv:1803.11175*, 2018.

[195] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "COVAREP—A collaborative voice analysis repository for speech technologies," in *2014 ieee international conference on acoustics, speech and signal processing (icassp)*, 2014, pp. 960–964.

[196] V. Vonikakis, Y. Yazici, V. D. Nguyen, and S. Winkler, "Group happiness assessment using geometric features and dataset balancing," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 2016, pp. 479–486.

[197] C. T. Sreeramareddy, K. v Panduru, J. Menten, and J. van den Ende, "Time delays in diagnosis of pulmonary tuberculosis: a systematic review of literature," *BMC infectious diseases*, vol. 9, no. 1, pp. 1–10, 2009.

[198] H. Rachman *et al.*, "Unique transcriptome signature of Mycobacterium tuberculosis in pulmonary tuberculosis," *Infection and immunity*, vol. 74, no. 2,

pp. 1233–1242, 2006.

[199] World Health Organization", “Global tuberculosis report 2018,” 2019.

[200] J. F. Ludvigsson, J. Wahlstrom, J. Grunewald, A. Ekbom, and S. M. Montgomery, “Coeliac disease and risk of tuberculosis: a population based cohort study,” *Thorax*, vol. 62, no. 1, pp. 23–28, 2007.

[201] S. S. Jick, E. S. Lieberman, M. U. Rahman, and H. K. Choi, “Glucocorticoid use, other associated factors, and the risk of tuberculosis,” *Arthritis Care & Research: Official Journal of the American College of Rheumatology*, vol. 55, no. 1, pp. 19–26, 2006.

[202] J. E. Oeltmann, J. S. Kammerer, E. S. Pevzner, and P. K. Moonan, “Tuberculosis and substance abuse in the United States, 1997-2006,” *Archives of Internal Medicine*, vol. 169, no. 2, pp. 189–197, 2009.

[203] C.-H. Lee *et al.*, “Risk factors for pulmonary tuberculosis in patients with chronic obstructive airway disease in Taiwan: a nationwide cohort study,” *BMC infectious diseases*, vol. 13, no. 1, pp. 1–11, 2013.

[204] S. E. Weinberger, B. A. Cockrill, and J. Mandel, *Principles of Pulmonary Medicine E-Book*. Elsevier Health Sciences, 2017.

[205] I. A. Campbell and O. Bah-Sow, “Pulmonary tuberculosis: diagnosis and treatment,” *Bmj*, vol. 332, no. 7551, pp. 1194–1197, 2006.

[206] H. Rée, “Treatment of tuberculosis: Guidelines for national programmes . D. Maher, P. Chaulet, S. Spinaci & A. Harries (writing committee). Geneva:

World Health Organization, 1997. 78pp. Price£ 7.50. WHO/TB/97.220." Royal Society of Tropical Medicine and Hygiene, 1999.

[207] J.-X. Chen *et al.*, "Novel therapeutic evaluation biomarkers of lipid metabolism targets in uncomplicated pulmonary tuberculosis patients," *Signal transduction and targeted therapy*, vol. 6, no. 1, pp. 1–11, 2021.

[208] G. M. B. Kussen, L. M. Dalla-Costa, A. Rossoni, and S. M. Raboni, "Interferon-gamma release assay versus tuberculin skin test for latent tuberculosis infection among HIV patients in Brazil," *Brazilian Journal of Infectious Diseases*, vol. 20, pp. 69–75, 2016.

[209] R. S. Wallis *et al.*, "Biomarkers and diagnostics for tuberculosis: progress, needs, and translation into practice," *The lancet*, vol. 375, no. 9729, pp. 1920–1937, 2010.

[210] A. S. Piatek *et al.*, "GeneXpert for TB diagnosis: planned and purposeful implementation," *Global Health: Science and Practice*, vol. 1, no. 1, pp. 18–23, 2013.

[211] I. Brock, K. Weldingh, T. Lillebaek, F. Follmann, and P. Andersen, "Comparison of tuberculin skin test and new specific blood test in tuberculosis contacts," *American journal of respiratory and critical care medicine*, vol. 170, no. 1, pp. 65–69, 2004.

[212] J. Nielsen and M. C. Jewett, *Metabolomics: a powerful tool in systems biology*, vol. 18. Springer Science & Business Media, 2007.

[213] E. D. Harris, “Biochemical facts behind the definition and properties of metabolites,” *Biochemistry and Biophysics and Faculty of Nutrition Texas A&M University*, 2017.

[214] X. Zhang, X. Zhu, C. Wang, H. Zhang, and Z. Cai, “Non-targeted and targeted metabolomics approaches to diagnosing lung cancer and predicting patient prognosis,” *Oncotarget*, vol. 7, no. 39, p. 63437, 2016.

[215] U. Vrhovsek *et al.*, “A versatile targeted metabolomics method for the rapid quantification of multiple classes of phenolics in fruits and beverages,” *Journal of agricultural and food chemistry*, vol. 60, no. 36, pp. 8831–8840, 2012.

[216] C. G. Adu-Gyamfi *et al.*, “Plasma indoleamine 2, 3-dioxygenase, a biomarker for tuberculosis in human immunodeficiency virus-infected patients,” *Clinical Infectious Diseases*, vol. 65, no. 8, pp. 1356–1363, 2017.

[217] Y. Suzuki *et al.*, “Serum indoleamine 2, 3-dioxygenase activity predicts prognosis of pulmonary tuberculosis,” *Clinical and Vaccine Immunology*, vol. 19, no. 3, pp. 436–442, 2012.

[218] Y. Cho *et al.*, “Identification of serum biomarkers for active pulmonary tuberculosis using a targeted metabolomics approach,” *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.

[219] J. K. Frediani *et al.*, “Plasma metabolomics in human pulmonary tuberculosis disease: a pilot study,” *PloS one*, vol. 9, no. 10, p. e108854, 2014.

[220] D. S. Wishart, “Computational approaches to metabolomics,”

Bioinformatics methods in clinical research, pp. 283–313, 2010.

[221] D. W. Zimmerman, "Comparative power of Student t test and Mann-Whitney U test for unequal sample sizes and variances," *The Journal of Experimental Education*, vol. 55, no. 3, pp. 171–174, 1987.

[222] G. Brandacher *et al.*, "Prognostic value of indoleamine 2, 3-dioxygenase expression in colorectal cancer: effect on tumor-infiltrating T cells.," *Clinical cancer research*, vol. 12, no. 4, pp. 1144–1151, 2006.

[223] S. v Schmidt and J. L. Schultze, "New insights into IDO biology in bacterial and viral infections," *Frontiers in immunology*, vol. 5, p. 384, 2014.

[224] A. W. S. Yeung, A. C. Terentis, N. J. C. King, and S. R. Thomas, "Role of indoleamine 2, 3-dioxygenase in health and disease," *Clinical science*, vol. 129, no. 7, pp. 601–672, 2015.

[225] D. H. Munn and A. L. Mellor, "Indoleamine 2, 3-dioxygenase and tumor-induced tolerance," *The Journal of clinical investigation*, vol. 117, no. 5, pp. 1147–1154, 2007.

[226] Y. Suzuki *et al.*, "Indoleamine 2, 3-dioxygenase in the pathogenesis of tuberculous pleurisy," *The International journal of tuberculosis and lung disease*, vol. 17, no. 11, pp. 1501–1506, 2013.

[227] Y. Luo *et al.*, "A combination of iron metabolism indexes and tuberculosis-specific antigen/phytohemagglutinin ratio for distinguishing active tuberculosis from latent tuberculosis infection," *International Journal of*

Infectious Diseases, vol. 97, pp. 190–196, 2020.

[228] Y. Li, M. Kuhn, A.-C. Gavin, and P. Bork, "Identification of metabolites from tandem mass spectra with a machine learning approach utilizing structural features," *Bioinformatics*, vol. 36, no. 4, pp. 1213–1218, 2020.

[229] C. Zhu, Z. Liu, Z. Li, S. Mei, and Z. Hu, "The performance and limitation of T-SPOT. TB for the diagnosis of TB in a high prevalence setting," *Journal of thoracic disease*, vol. 6, no. 6, p. 713, 2014.

[230] W. Wang, W. Li, N. Ma, and G. Steinhoff, "Non-viral gene delivery methods," *Current pharmaceutical biotechnology*, vol. 14, no. 1, pp. 46–60, 2013.

[231] J. A. Zuris *et al.*, "Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo," *Nature biotechnology*, vol. 33, no. 1, pp. 73–80, 2015.

[232] J. Dobson, "Gene therapy progress and prospects: magnetic nanoparticle-based gene delivery," *Gene therapy*, vol. 13, no. 4, pp. 283–287, 2006.

[233] S. v Vinogradov, T. K. Bronich, and A. v Kabanov, "Nanosized cationic hydrogels for drug delivery: preparation, properties and interactions with cells," *Advanced drug delivery reviews*, vol. 54, no. 1, pp. 135–147, 2002.

[234] A. H. Nguyen *et al.*, "Polyethylenimine-coated iron oxide magnetic nanoparticles for high efficient gene delivery," *Applied Nanoscience*, vol. 8, no. 4, pp. 811–821, 2018.

[235] E. Neumann, M. Schaefer-Ridder, Y. Wang, and P. Hofschneider, “Gene transfer into mouse lyoma cells by electroporation in high electric fields.,” *The EMBO journal*, vol. 1, no. 7, pp. 841–845, 1982.

[236] M. P. Stewart, A. Sharei, X. Ding, G. Sahay, R. Langer, and K. F. Jensen, “In vitro and ex vivo strategies for intracellular delivery,” *Nature*, vol. 538, no. 7624, pp. 183–192, 2016.

[237] V. K. Bajpai *et al.*, “A sustainable graphene aerogel capable of the adsorptive elimination of biogenic amines and bacteria from soy sauce and highly efficient cell proliferation,” *ACS applied materials & interfaces*, vol. 11, no. 47, pp. 43949–43963, 2019.

[238] S. Shukla *et al.*, “Sustainable graphene aerogel as an ecofriendly cell growth promoter and highly efficient adsorbent for histamine from red wine,” *ACS applied materials & interfaces*, vol. 11, no. 20, pp. 18165–18177, 2019.

[239] D. Huang *et al.*, “Impact of low-intensity pulsed ultrasound on transcription and metabolite compositions in proliferation and functionalization of human adipose-derived mesenchymal stromal cells,” *Scientific reports*, vol. 10, no. 1, pp. 1–19, 2020.

[240] Z. Li *et al.*, “Nanoparticle depots for controlled and sustained gene delivery,” *Journal of Controlled Release*, vol. 322, pp. 622–631, 2020.

[241] I. Villate-Beitia *et al.*, “Non-viral vectors based on magnetoplexes, lipoplexes and polyplexes for VEGF gene delivery into central nervous system

cells," *International journal of pharmaceutics*, vol. 521, no. 1–2, pp. 130–140, 2017.

[242] S. Rahmani *et al.*, "Novel chitosan based nanoparticles as gene delivery systems to cancerous and noncancerous cells," *International journal of pharmaceutics*, vol. 560, pp. 306–314, 2019.

[243] W. Jin *et al.*, "Transfection of difficult-to-transfect rat primary cortical neurons with magnetic nanoparticles," *Journal of biomedical nanotechnology*, vol. 14, no. 9, pp. 1654–1664, 2018.

[244] J. P. Yang and L. Huang, "Overcoming the inhibitory effect of serum on lipofection by increasing the charge ratio of cationic liposome to DNA," *Gene therapy*, vol. 4, no. 9, pp. 950–960, 1997.

Appendix

Table A. 1 Comparison Between Different Classifiers with Different Dimensionality Reduction methods on the DAIC-WOZ Development Set (The unimodal models shown in bold achieved the best performance).

Features	Model	F1 score (Healthy)	F1 score (Depressed)	F1 score (Weighted)
MFCCs+ COVAREP	SVM	0.783	0.583	0.714
	PCA + SVM	**0.826**	**0.667**	**0.771**
	XGBoost + SVM	0.486	0.424	0.465
	XGBoost	0.652	0.333	0.543
	PCA + XGBoost	0.622	0.320	0.514
	XGBoost + XGBoost	0.739	0.500	0.657
	KNN	0.526	0.438	0.486
	PCA + KNN	0.578	0.240	0.457
	XGBoost + KNN	0.682	00462	0.606
FAUs	SVM	0.651	0.444	0.580
	PCA + SVM	0.619	0.429	0.554
	XGBoost + SVM	0.711	0.480	0.632
	XGBoost	0.792	0.545	0.707
	PCA + XGBoost	0.694	0.286	0.554
	XGBoost + XGBoost	**0.833**	**0.636**	**0.766**
	KNN	0.622	0.320	0.519
	PCA + KNN	0.652	0.333	0.543
	XGBoost + KNN	0.638	0.261	0.509

Table A. 2 Correlation between different metabolites and age in different groups.

Correlation	Kyn	Trp	IDO
Age	0.33	0.05	0.26

Table A. 3 Performance of logistic regression models with various biomarkers for discriminating different groups along with the hypothesis test results.

	HC vs NTB					
	Kyn		Trp		IDO	
	Discovery	Validation	Discovery	Validation	Discovery	Validation
AUC	0.83 (+/- 0.10)	0.87	0.62 (+/- 0.07)	0.63	0.76 (+/- 0.08)	0.68
Accuracy	0.71 (+/- 0.10)	0.75	0.61 (+/- 0.08)	0.71	0.67 (+/- 0.09)	0.58
Specificity	0.74 (+/- 0.09)	0.69	0.71 (+/- 0.07)	0.69	0.76 (+/- 0.10)	0.62
Sensitivity	0.69 (+/- 0.15)	0.82	0.51 (+/- 0.10)	0.73	0.57 (+/- 0.11)	0.55
	Mann-Whitney *U* test	*t*-test	Mann-Whitney *U* test	*t*-test	Mann-Whitney *U* test	*t*-test
P-value	8.68E-8	2.01E-7	0.021	0.060	2.70E-5	0.00012
	HC vs LTBI					
	Kyn		Trp		IDO	
	Discovery	Validation	Discovery	Validation	Discovery	Validation
AUC	0.61 (+/- 0.14)	0.65	0.72 (+/- 0.13)	0.89	0.68 (+/- 0.12)	0.72
Accuracy	0.58 (+/- 0.07)	0.64	0.66 (+/- 0.09)	0.76	0.59 (+/- 0.09)	0.68
Specificity	0.54 (+/- 0.11)	0.54	0.56 (+/- 0.19)	0.69	0.64 (+/- 0.13)	077
Sensitivity	0.61 (+/- 0.16)	0.75	0.75 (+/- 0.15)	0.83	0.54 (+/- 0.11)	0.58
	Mann-Whitney *U* test	*t*-test	Mann-Whitney *U* test	*t*-test	Mann-Whitney *U* test	*t*-test
P-value	0.022	0.075	7.27E-6	1.22E-5	0.00096	0.0028

	NTB vs LTBI					
	Kyn		**Trp**		**IDO**	
	Discovery	**Validation**	**Discovery**	**Validation**	**Discovery**	**Validation**
AUC	0.92 (+/- 0.06)	0.79	0.86 (+/- 0.13)	0.81	0.58 (+/- 0.21)	0.51
Accuracy	0.80 (+/- 0.05)	0.75	0.78 (+/- 0.11)	0.67	0.58 (+/- 0.08)	0.50
Specificity	0.77 (+/- 0.13)	0.75	0.73 (+/- 0.06)	0.50	0.42 (+/- 0.12)	0.50
Sensitivity	0.84 (+/- 0.09)	0.75	0.82 (+/- 0.20)	0.83	0.73 (+/- 0.14)	0.50
	Mann-Whitney *U* test	***t*-test**	**Mann-Whitney *U* test**	***t*-test**	**Mann-Whitney *U* test**	***t*-test**
P-value	5.63E-11	4.83E-10	1.48E-8	4.66E-9	0.082	0.18
	Control vs NTB					
	Kyn		**Trp**		**IDO**	
	Discovery	**Validation**	**Discovery**	**Validation**	**Discovery**	**Validation**
AUC	0.83 (+/- 0.09)	0.92	0.69 (+/- 0.06)	0.76	0.63 (+/- 0.11)	0.77
Accuracy	0.76 (+/- 0.12)	0.81	0.69 (+/- 0.03)	0.70	0.68 (+/- 0.03)	0.70
Specificity	0.92 (+/- 0.09)	0.92	0.92 (+/- 0.04)	0.96	0.96 (+/- 0.03)	1.00
Sensitivity	0.41 (+/- 0.24)	0.58	0.21 (+/- 0.06)	0.17	0.06 (+/- 0.06)	0.08
	Mann-Whitney *U* test	***t*-test**	**Mann-Whitney *U* test**	***t*-test**	**Mann-Whitney *U* test**	***t*-test**
P-value	9.86E-12	1.96E-09	7.13E-6	2.42E-5	0.00086	0.0064

Table A. 4 Performance of logistic regression models for discriminating different binary groups.

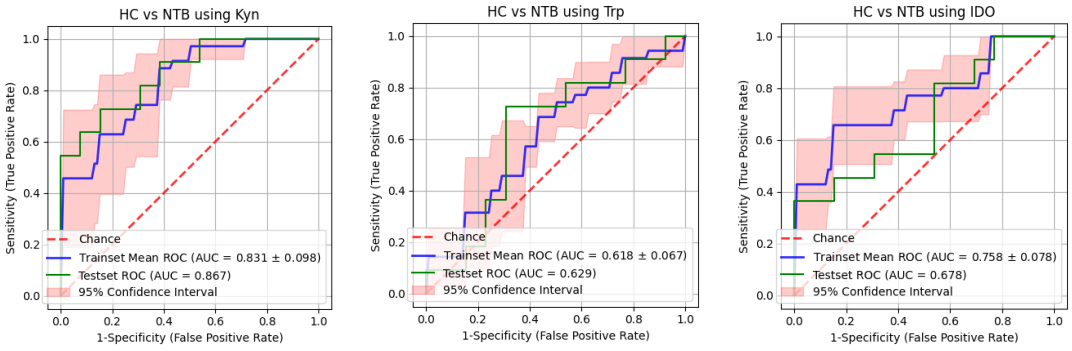
	HC vs NTB		**HC vs LTBI**	
	Discovery	**Validation**	**Discovery**	**Validation**
AUC	0.83 (+/- 0.10)	0.82	0.71 (+/- 0.10)	0.88
Accuracy	0.75 (+/- 0.13)	0.75	0.67 (+/- 0.11)	0.76

Specificity	0.82 (+/- 0.14)	0.77	0.62 (+/- 0.15)	0.69
Sensitivity	0.69 (+/- 0.15)	0.73	0.72 (+/- 0.14)	0.83

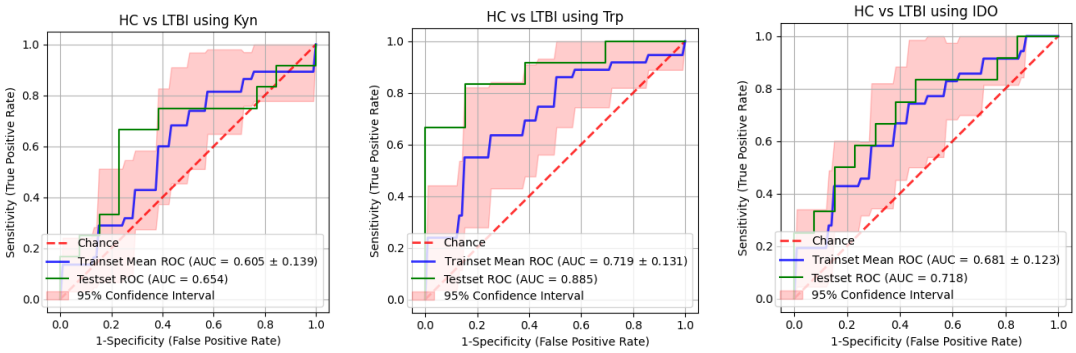
	NTB vs LTBI		**Control vs NTB**	
	Discovery	**Validation**	**Discovery**	**Validation**
AUC	0.89 (+/- 0.08)	0.83	0.82 (+/- 0.09)	0.92
Accuracy	0.78 (+/- 0.07)	0.79	0.75 (+/- 0.12)	0.78
Specificity	0.74 (+/- 0.09)	0.83	0.91 (+/- 0.09)	0.92
Sensitivity	0.82 (+/- 0.17)	0.75	0.41 (+/- 0.24)	0.50

Table A. 5 Performance of logistic regression model for discriminating ATB vs NTB.

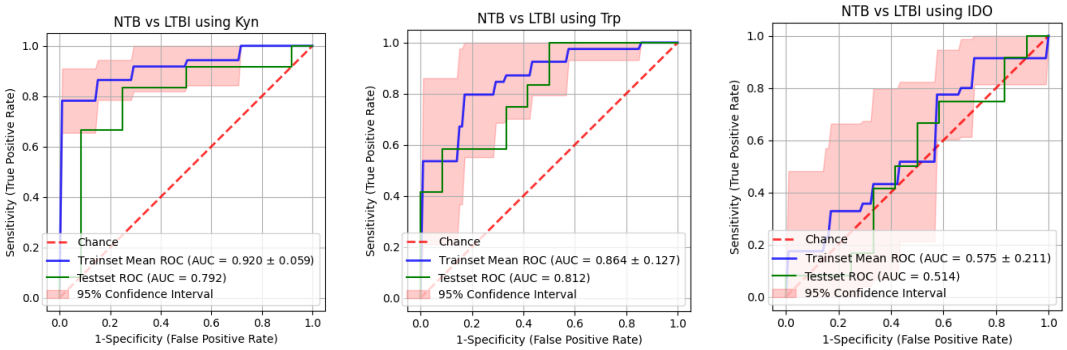
	Using t-spot		
	Discovery	**Validation**	**External Validation**
AUC	0.83 (+/- 0.00)	0.82	0.77
Accuracy	0.82 (+/- 0.09)	0.80	0.78
Specificity	0.75 (+/- 0.20)	0.64	0.70
Sensitivity	0.92 (+/- 0.09)	1.00	0.83



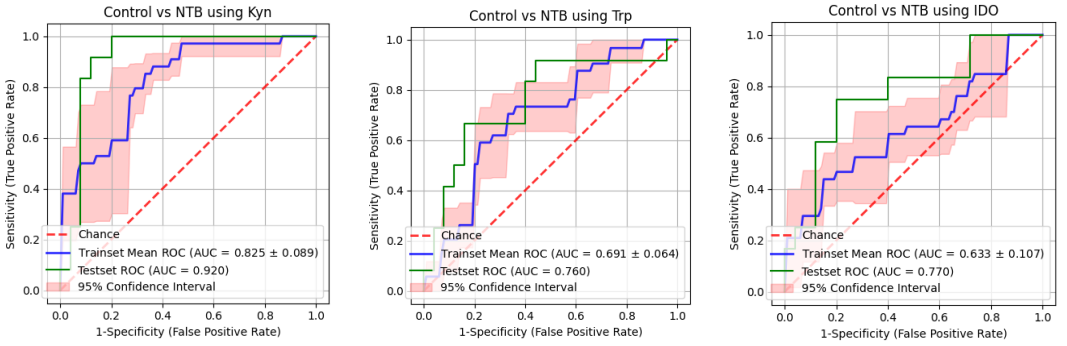
(a) (b) (c)



(d) (e) (f)



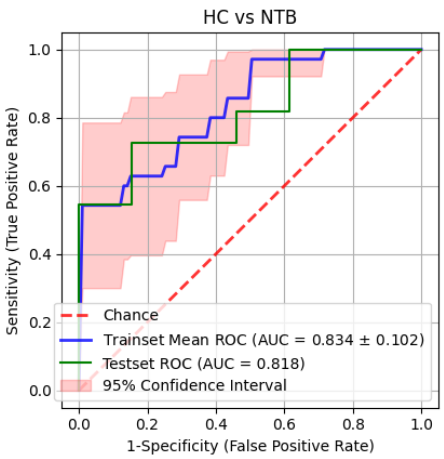
(g) (h) (i)



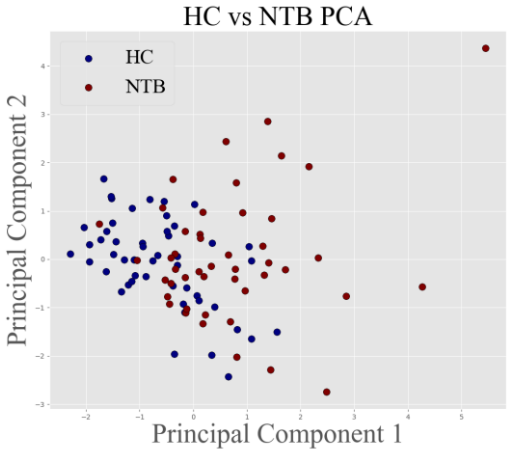
(j) (k) (l)

Figure A. 1 Receiver-operating characteristic (ROC) curves of the logistic regression model; (a) using Kyn for discriminating HC and NTB patients ; (b) using Trp for discriminating HC and NTB patients; (c) using IDO for discriminating HC and NTB patients; (d) using Kyn for discriminating HC and LTBI patients; (e) using Trp for discriminating HC and LTBI patients; (f) using IDO for discriminating HC and LTBI patients; (g) using Kyn for discriminating NTB and LTBI patients; (h) using Trp for

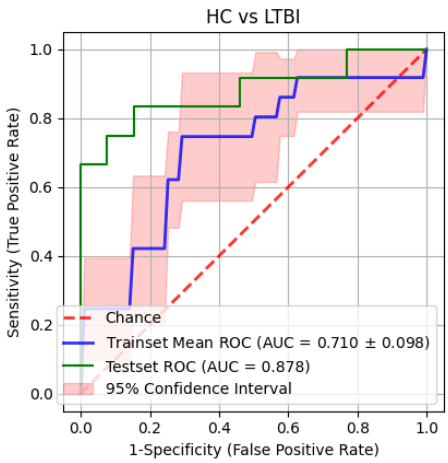
discriminating NTB and LTBI patients; (i) using IDO for discriminating NTB and LTBI patients; (j) using Kyn for discriminating control and NTB patients; (j) using Trp for discriminating control and NTB patients; (l) using IDO for discriminating control and NTB patients. The ROC curve is plotted by the true positive rate (TPR) against the false positive rate (FPR) at different thresholds. ROC curves with 95% confidence interval of these logistic regression models are shown for distinguishing among HC, LTBI, and NTB utilizing Kyn, Trp, and IDO separately. The blue curve is the mean ROC, and the red regions show the 95% confidence intervals in the discovery set over five folds. The green curve indicates the ROC curve on the validation set. The best classification will create a point at coordinates (0,1), representing 100% sensitivity and 100% specificity.



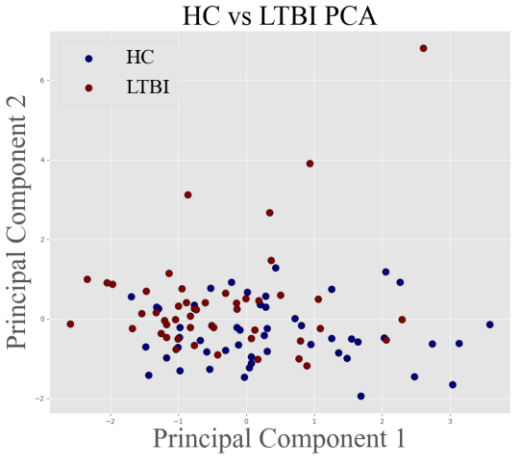
(a)



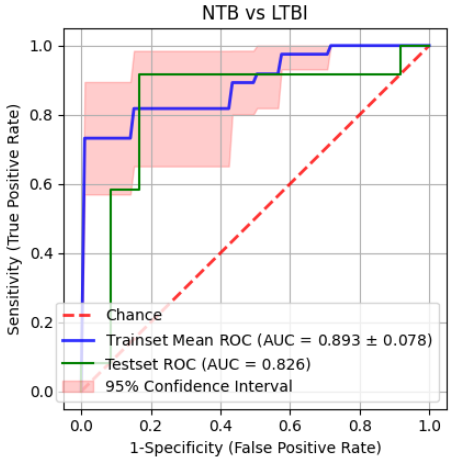
(b)



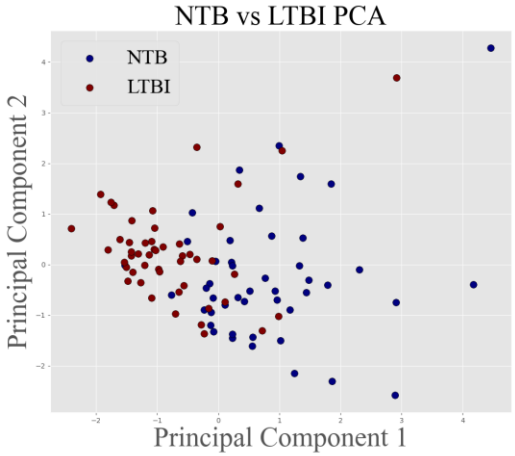
(c)



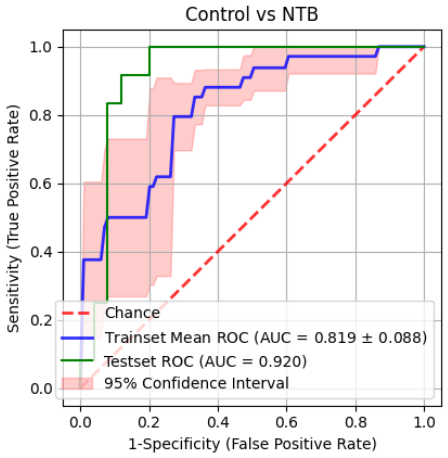
(d)



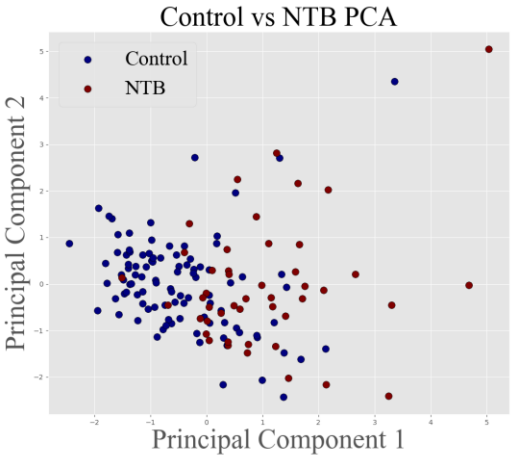
(e)



(f)



(g)



(h)

Figure A. 2 ROC curves of the logistic regression model using Kyn, Trp, and IDO:

(a) discriminating HC and NTB patients; (c) discriminating HC and LTBI patients; (e) discriminating NTB and LTBI patients; (g) discriminating control and NTB patients. PCA plot shows the ability to discriminate different groups: (b) discriminating HC and NTB patients; (d) discriminating HC and LTBI patients; (f) discriminating NTB and LTBI patients; (h) discriminating control and NTB patients. ROC curves with 95% confidence interval of these logistic regression models using the biomarkers together were performed to visualize the performance of the classification model. Principal Component Analysis with the data from different combined groups was performed and visualized the first two components, which can show the ability to distinguish different groups.

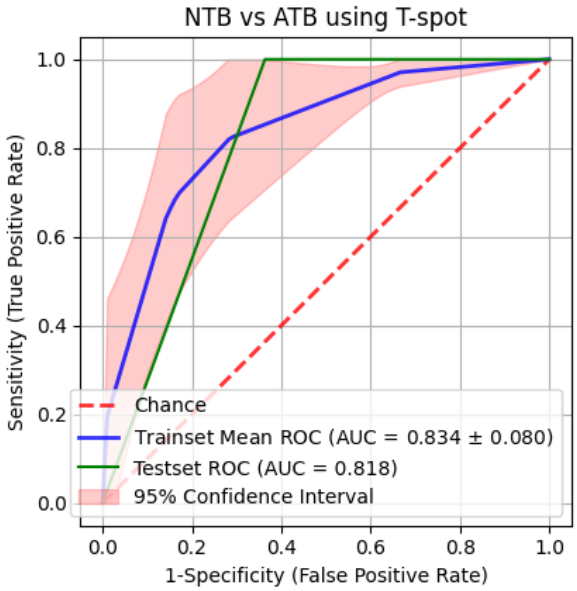


Figure A. 3 ROC curves of the logistic regression model for discriminating NTB and ATB patients just using t-spot. ROC curves with a 95% confidence interval were

employed to evaluate the predictive value of the t-spot in classifying NTB and ATB. The t-spot cannot predict ATB accurately.