Evaluating Image Training Systems for Medical Students

by

Reza Sobhannejad

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

Master of Science

Department of Computing Science

University of Alberta

Abstract

Skin cancer is one of the common and most fatal cancers. Therefore, it is important to be able to diagnose skin lesions and detect this cancer before it is too late.

Learning to distinguish between sick and healthy lesions is key. However, there are two levels in which one can learn to distinguish between malignant (sick skin lesions) and benign (healthy skin lesions) cases and knowing which one is more effective is not a trivial task.

The first learning level is known as basic level and consists of distinguishing between benign and malignant cases. The second level is the subordinate level and focuses on recognizing the subgroups of malignant and benign cases. In this work, we design an experiment to assess and compare the two learning levels, basic versus subordinate level learning. The results show that the participants of the basic level were better able to distinguish healthy from sick cases compared to participants of the subordinate level.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the 21st century, the daily life is going towards using electronic devices. In the past decade, we have seen many applications that help us do what is in our mind. A part of these applications are the ones that are helping people learn different subjects. For example, the applications for learning different languages and many of these applications have succeeded in doing so. In this project, our goal is to develop an application which can help medical students identify different skin lesions and separate the malignant (harmful) group from benign (healthy).

1.1 Motivation

skin cancer is the most common form of cancer [12]. This fact makes it interesting to spend some time on this matter and think about the ways that we can help patients fight against cancer. In this section, we will mention some of the reasons that make the topic even more interesting.

1.1.1 Wrong Identification

Many skin cancer patients are suffering because the identification of the lesion has not been successful. They do not go to doctor or pay enough attention to the lesion until it is too late and sometimes nothing can be done in that step. So the doctor identifying the lesion correctly is critical for the patient.

You might think that the doctor can report the lesion as malignant unless he is absolutely sure that the lesion is benign. It seems to be a good solution.

No one will have the problem that was just mentioned. The problem here is that when the doctor reports the lesion as malignant, more tests are required in order to find necessary details about the lesion, so the patient needs to see a dermatologist. Besides being so expensive, patients usually need to wait for about 6 months sometimes and this causes anxiety. That is why is step is preferred to be avoided unless necessary. For this reason, it seems to be better to increase the identification's accuracy instead of reporting every suspicious lesion as malignant.

The problem is that medical students do not see enough skin lesions in their education, so they do not have the required experience for identifying malignant lesions. That is why there is a set of rules called ABCD, that help the doctors identify the skin lesions. In the system proposed in this thesis, this problem is solved by showing the students adequate skin lesion images.

1.1.2 ABCD rules

The most popular method for identifying the skin lesions is called ABCD rules [4]. Currently, students are taught to look for a set of visual features indicative of skin cancer. These rules are:

- **A: Asymmetry**

  One half of the mole does not match the other half.

- **B: Border Irregularity**

  The edges of the mole are irregular, ragged, blurred, or notched.

- **C: Colour**

  The color over the mole is not the same. There may be differing shades of tan, brown, or black and sometimes patches of red, blue or white.

- **D: Diameter**

  The mole is larger than 6 millimeters(about 1/4 inch or about the size of a pencil eraser) or is growing larger.

There is also another rule which was added to the set later: **E** for **Evolving**. This rule classifies the evolving lesions as malignant [1].

Unfortunately, the empirical evidence supporting the relationship between the presence of these features and skin cancer is weak. Besides, what happens if a lesion has light colour but sharp edges? There is no such rule that can be used when the lesion is classified as benign with one rule and malignant with another one. In this case, you will have to find another way for classifying the lesion.

1.1.3 Experience

Whenever we are sick and we want to go to a doctor, one of the factors that help us choose the doctor is his experience in the subject. Why? We all believe that experience is a really important factor in how good the doctor (or someone in any other job) is. The problem is that the experience has been achieved over time and it is not an element that can be transferred to new doctors so easily.

For example, when you have a problem like the one in the previous section, experience seems to be the only solution. You can not classify the lesion unless you have enough experience. In this thesis I am trying to develop an application that lets medical students learn skin lesions based on experts' experience.

1.2 Problem Definition

Our goal in this thesis is to help doctors identify malignant lesions correctly. In order to succeed in this task we need to find the best way for teaching them. In this section the categorization of skin lesions will be introduced and then we will talk about the different possible levels of learning skin lesions.

Skin lesions have two main categories of skin lesions: benign and malignant. Since there are different types of lesions in each category, there are four perceptual subcategories for each one. You can see this subcategories in Figure 1.1.

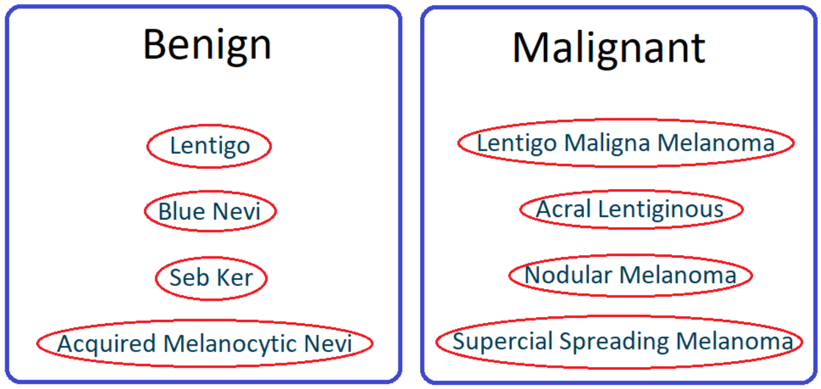


Figure 1.1: Skin lesion categories and subcategories

Each skin lesion can be placed in one of the boxes as well as one of the eight ellipses. The first labeling is called basic level and the second is called subordinate level. The final goal of this project is to help doctors to classify the lesions in the basic level, because if you have a malignant lesion it does not matter in which subcategory it is. You have to go to a dermatologist anyway.

There is a research on learning birds that shows learning in the subordinate level leads to a better generalization for new birds [11]. This rises a question in our subject: which one of the basic level learning and the subordinate level learning is more effective in learning skin lesions. In this thesis we are trying to find the answer to this question.

1.3 Methodology

For addressing this problem we need five different components:

- **Data collector:** In the first step we collect experts' opinion on different lesions. We have an Android and an iOS application which let experts help us gather these lesions. In this application, the experts can take a picture of the lesion and their comment on the lesion and send them to our server at the University of Alberta.

- **The server:** We also have to have a server which listens to these kinds of requests and stores these pictures in the database that exists on the server.

- **Data Converter:** Except the picture of the lesion, the only data that we have in the database is an image from the expert's comments on the lesion. In order to be able to work with these images, we need to have an intern labeling the lesion with the right group and subgroup. Besides, the images that are stored in the server are not useful necessarily (e.g. The image might not be clear enough). This kind of inputs will be eliminated in the same step.

- **Teaching application:** Finally we can use the processed images in the previous step. Here we have an Android application which uses these lesions and teaches medical students the difference between benign and malignant (or the differences between the subtypes).

- **Pretest And Post-test:** This part is supposed to make sure that the participants do not know anything about the skin lesions at the beginning. It also measures their knowledge before and after the experiment which shows if they have learned something.

Using these tools, an experiment has been designed to compare the effectiveness of the basic level learning to the subordinate level learning.

1.4 Thesis Statement

In this study, I have tried to follow some ideas from the education studies and create our application based on them. Besides, we tried to develop an application which can be used not only for an experiment but also for learning lesions regularly. We claim that:

For distinguishing the difference between benign and malignant, learning skin lesions in the basic level is more effective than learning them in the subordinate level and we can develop a system to show this.

1.5 Thesis Contribution

This project had been started before I started my thesis. For the data collection part, an Android application had been written by Farrukh Ahmed. In my thesis, we developed an iOS application similar to what we had in Android.

In the second step, I implemented the server in a way that it was able to connect to both Android and iOS applications. The server saves the pictures in a database in the server, so we can use them in the next step.

Then, we developed an Android application for the teaching part based on a combination of different structures that showed to be effective. I will talk more about the structure in the *System Design* chapter. This application consists of different teaching and assessment sessions for both the basic level and the subordinate level.

At last, we developed another Android application for the pretest and the post-test. This application is a single session test which participants take before and after the learning process. These exams need to be the same but there is no problem with that, because the participants do not see the correct label of each image in either of the tests and the images are completely different than the images used in the learning application.

1.6 Organization of Dissertation

For describing the thesis I will have three more chapters. The order is as follows:

- **Chapter 2:** Related Work

In this chapter, I will review some research on learning methods in different areas. The main focus will be on birds and then rocks.

- **Chapter 3:** System Design

  In this chapter, I will explain the details of the data collector, the teaching application, the pretest application and the server. Then the connection between the components will be explained.

- **Chapter 4:** Experiment

  This chapter will include the procedure and the results of our experiment using the system introduced in chapter 3.

- **Chapter 5:** Conclusion and Future Work

  This will be a brief review of the thesis and some possible paths for continuing this project.

Chapter 2

Related Work

The goal of this project is to compare the basic level learning and the subordinate level learning. We are looking for the best level of learning which transfers knowledge and expertise from experts to the younger generation. There have been some studies comparing these two levels of learning. This question has been asked about birds and rocks. In this chapter I will explain these studies and we will see the results.

2.1 Experience

Transferring expertise to the younger generation has been so important that research shows that companies which have been able to transfer the experience and knowledge effectively have been more successful than the companies which have not [2]. The question then arises as to what is experience exactly?

There is another study on training greeble (a novel class of objects) experts which indicates that expertise is neither described by a single term nor evaluated by a single task [5]. This means we do not know what makes some people expert really. What should we do exactly when we want to transfer the expertise? The answer to this question depends on the subject. The features of an expert can be different in two separate areas. In one field the most important feature of the expert may be his accuracy while in another, it may be his speed. In our subject (skin lesions) experts are able to identify malignant lesions with high accuracy. In the next section we will see which one of the basic level learning and the subordinate level learning has been more effective in other subjects.

2.2 Visual Category Learning

In this thesis, we are focusing on teaching younger students about skin cancer. In this problem, their task will be to look at the skin lesion and classify the lesion as benign or malignant. We know that visual category learning connects specific perceptual experience with abstract conceptual knowledge [8]. Now in this chapter, we will overview some of the research on visual category learning in different areas. Then in Chapter 3 we will talk about the same problem in skin lesions.

2.2.1 Learning Birds

In 2016, a team from the University of Colorado, University of Victoria and Brown University, designed an experiment to find more details on this subject [11]. Their experiment was on two different types of birds: owls and wading birds. They had two different groups for their experiment. One of them studied owls in the basic level and wading birds in the subordinate level and the other group studied wading birds in the basic level and owls in the subordinate level.

A pretest had been performed in order to make sure none of the participants had knowledge on these two different kinds of birds, so we know the comparison of their performance before and after the experiment is meaningful.

The experiment has three training tasks. The first one is called *naming*. In this task, participants see an image with a label: either the subordinate level label (e.g. "This is species Y") or the basic level label (e.g. "This is the Other type"). Then they will see an image and they should provide the right name.

The second task is called *category verification*. In this part, the participants will see a bird label (the subordinate level for their type and the basic level for the other), then a fixation point of 500 milliseconds and then they see an image. After that, they need to say if they think the label and the image match. They have two choices. There are two buttons of SAME and DIFFERENT and they need to select one of them.

The last training task is called *matching*. In this task, the participants are presented with a label first (the subordinate level for their type and the basic level for the other). Then there is a fixation point for 500 millisecond and after that, they see two side-by-side images. They should choose the image which pairs with the label.

After the experiment is done there will be a post-test. This test is designed to see if the training has been successful. In order to measure the success rate of the experiment, there are three groups of images in the post-test. The first group is the trained exemplars. The participants' performance in this group shows how much they have learned the images they have seen. In other words, this shows the level of their effort. The second group is the new exemplars of the same types they learned and this one shows their knowledge of the types they learned. The last, is new exemplars of the new types which shows their ability to generalize their knowledge. These three groups in the post-test help us understand if the training sessions have been successful and show us the direction in which the experiment has changed their ability. Below you can find some of the examples from these categories.

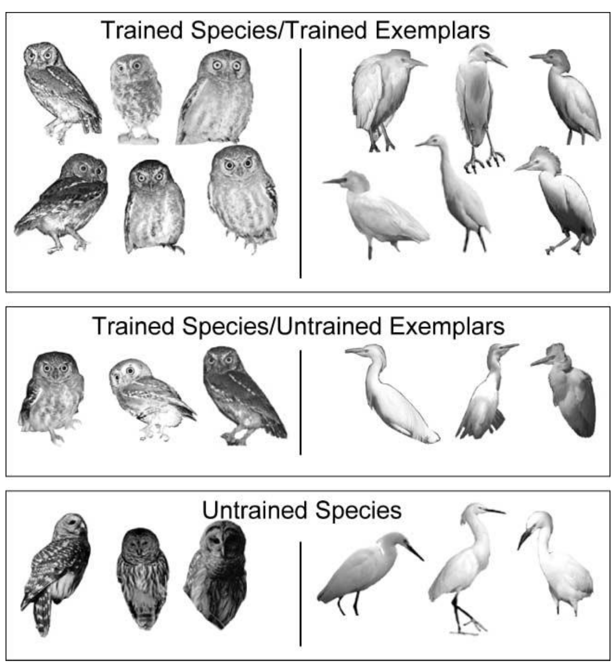


Figure 2.1: Different groups of images from the post-test. [11]

The results of this experiment which were obtained using ERP, shows that the part of brain which gets activated for the basic level is different than the part of brain which is amplified for the subordinate level (The part relate to the subordinate level is the same part related to face recognition which was found in another research [3]). This shows that the the basic level learning and the subordinate level learning are so different. The question is that which one of them is more effective.

The results show that each group has a better generalization for the type of birds which they learned in the subordinate level. They have been able to identify new images of the new types of the group of images which they learned in the subordinate level. This means that in this subject (birds) learning in the subordinate level leads to a better generalization compared to learning in the basic level. Although, this does not mean that if you learn birds in the subordinate level, you can deal with new bird images just as easily as you deal with the old ones. In fact there is another experiment about brain activities for when we look at different faces and the results show that we have a stronger brain activity when we see familiar faces compared to when we see new faces [10].

2.2.2 Learning Rocks

In 2016, a team from Indiana University and Washington University designed an experiment with a similar subject [7]. In the first step, they introduce family-resemblance principle from psychology which states: "members of the same category share bundles of characteristic features that are not shared by members of contrasting categories". Then they defined two types of structure: compact structure and dispersed structure. The compact structure reflects the classic assumption of the family-resemblance principle, but the dispersed structure is exactly the opposite. Figure 2.3 explains these two structures better.

Based on these definitions, rocks are more compatible with the dispersed structure, so they violate the family-resemblance principle. Rocks have three different types: igneous, metamorphic and sedimentary, where each of them

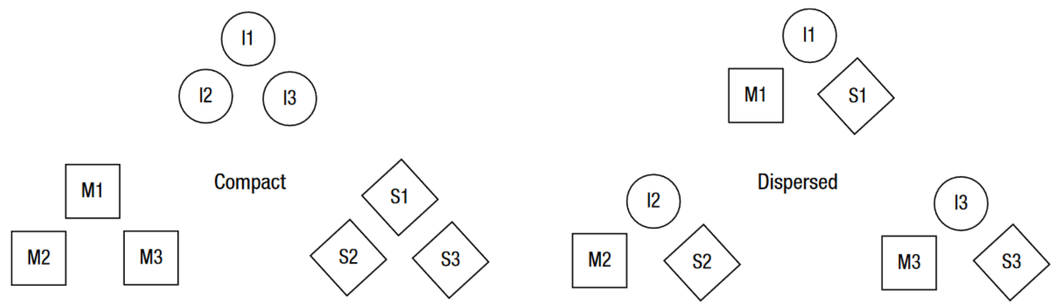


Figure 2.2: An example of compact structure vs dispersed structure [7]

has its own subtypes. So far there has been evidence that effective techniques for learning the dispersed category may be the exact opposite of the ones that are useful for learning the compact categories. Now, if we accept that rocks belong to the dispersed group, the question becomes which of the methods of the basic level learning and the subordinate level learning works better for them.

In order to find the answer to the question, first, there has to be an experiment for verifying that rocks really belong to the dispersed group. In this experiment, 30 subtypes of rocks have been chosen (10 subtypes of each type). Using images of these 30 subtypes, subjects provided pairwise similarity for all of the possible pairs. The chosen subtypes for this experiment are in Table 2.1.

Igneous	Metamorphic	Sedimentary
Andesite	Amphibolite	Bituminous coal
Basalt	Anthracite	Breccia
Diorite	Gneiss	Chert
Gabbro	Hornfels	Conglomerate
Granite	Marble	Dolomite
Obsidian	Migmatite	Micrite
Pegmatite	Phyllite	Rock gypsum
Peridotite	Quartzite	Rock salt
Pumice	Schist	Sandstone
Rhyolite	Slate	Shale

Table 2.1: Subtypes of Igneous, Metamorphic and Sedimentary [7]

Using these pairwise similarities, an average pairwise similarity was computed for each pair. These similarities were then passed to a standard nonmetric MDS analysis which locates each of the images in an M-dimensional space in a way that the space between every two rocks is as close to their average similarity as possible. This M-dimensional space shows that rocks from the same type are not really close together which means they are not that similar. This proves that rocks do not follow the family-resemblance principle. For the purpose of finding the factors of visual similarity some of the dimensions have been studied separately.

The first dimension seems to be showing the lightness of the rock, with the lighter ones on one side and the darker rocks on the other side. The second dimension is the average grain size. Thus, rocks with highly fragmented grains on the bottom and rocks without visible grains on top of Figure 2.4. The third dimension shows the coloration. That means the white and black rocks are on one side and colourful ones on the other side. And last, the fourth dimension seems to have a relation with the organization of the rocks. On the top of Figure 2.5, rocks usually have organized layers and on the bottom, rocks are more unorganized. There are two figures below that show some of the rocks in the M-dimensional space. All of the four dimensions can be seen in the figures below.

An important point about the types and subtypes of the rocks is that these categories have been made based on the manner of the rocks not on how they look. This can be one of the reasons that put the rocks in a dispersed structure category.

Now the question arises as to what happens if we choose different features and dimensions? Is it still going to be a dispersed structure? The answer is maybe, but the point is that it really seems unlikely that a new set of features or dimensions can change an extremely dispersed structure to a compact structure. We might be able to reduce the complexity of the structure by choosing another set of features, but having a compact structure for rocks seems very unlikely.

Now that we know the category structure of rocks is dispersed, another question can be asked. Is it better to learn rocks in the basic level or in the subordinate level? The classic research shows that we usually have more

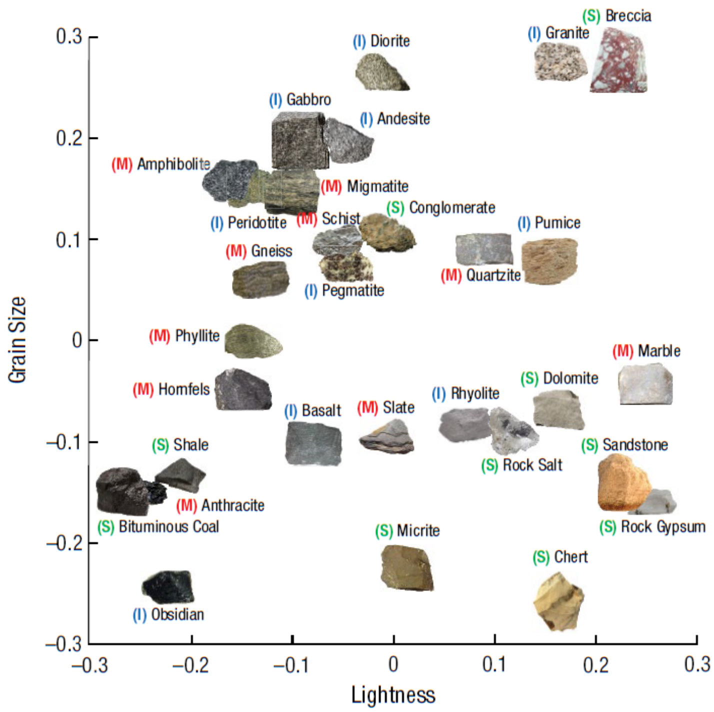


Figure 2.3: The dimensions of Lightness and Grain Size [7]

accurate classification when we learn something at a lower level (subordinate level) [6]. Is that the case for rocks too?

In the second experiment of this subject, a classification learning study was designed. In this experiment, we have three subtypes of each rock type and the participants are divided into two groups. One of them studies the rocks in the basic level and the other studies them in the subordinate level. These subtypes have been chosen in a way that the 9 subtypes form a compact structure. Similar experiment was performed with another 9 subtypes that form a dispersed structure.

For measuring the performance and the improvement there are a pretest and a post-test similar to the birds' experiment mentioned previously. The results of this experiment show that the subordinate level learning leads to a better performance in the study with the compact structure, but in the experiment with the dispersed structure, the basic level learning is better than the subordinate level learning. More importantly, the subordinate level learning

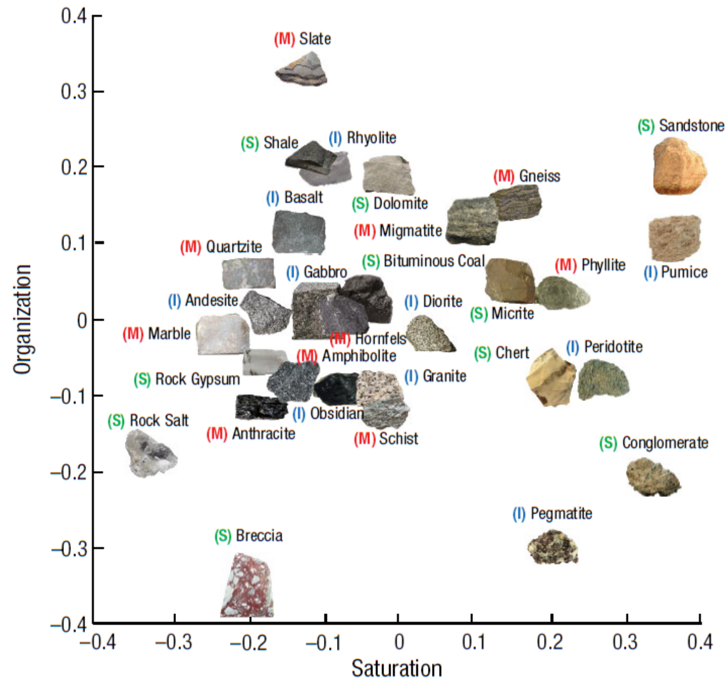


Figure 2.4: The dimensions of Saturation and Organization [7]

in conjunction with the basic level learning is better than each of them alone. This observation suggests that the best level of learning depends on whether the category structure is compact or dispersed.

2.3 Review and Conclusion

In this chapter, we reviewed some of the research about how to get an expert in visual category learning in different areas. We saw some experiments that were designed to find the best level of learning in order to get better at categorizing the objects. First, we found that once we become an expert in a subject, a specific part of the brain is activated whenever we see objects from the same group (N170) and then we saw that when learning a subject in the subordinate level, another part of the brain is amplified (N250). This fact brings about another question: if different parts of brain are amplified for different levels of learning, which level leads to a better performance?

We reviewed another experiment which suggests that the best learning level depends on whether the group structure is compact or dispersed. It shows that the combination of the basic level and the subordinate level learning is better for dispersed category structure and the basic level learning alone is the best way to learn a subject with compact category structure. In the next chapter, I will talk about the system design, used in our proposed method for learning the skin lesions.

Chapter 3

System Design

We explain in this chapter, the details of our system for teaching skin lesions to medical students. As we mentioned earlier there are three applications that help the students learn the lesions. The first application is the *Data Collector*. Using this application, dermatologists can take a picture of a skin lesion whenever they have a patient and send the image to a server, so we use them in the teaching application.

The main piece of the thesis is the second part which is the *teaching application*. In order to teach the students, this part shows them some images from different categories of skin lesions in different settings. This application is a combination of teaching sessions and assessments which will be explained later in this chapter.

The next part is the server which connects all of the pieces. This application receives the images sent from the data collector and saves them in the database. It also receives requests from the teaching application and sends the requested image back to the application. Without this piece, none of the two applications would work properly.

There is also another Android application for *pretest* and *post-test* which is not necessary for teaching. This application has been developed only to do the experiment. This is the main tool for measuring the students' ability to classify the lesions. The system design is as follows:

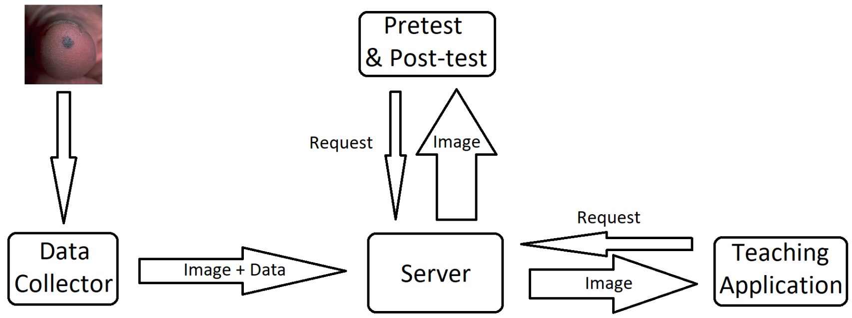


Figure 3.1: The system design

3.1 Data Collector

In order to teach medical students, we need to have a database with sufficient images of skin lesions. Therefore, the first step becomes gathering a database. In this task, we require dermatologists' assistance as they have access to real cases of skin lesions. Thus we need a simple application that helps dermatologists send skin lesion images to our database.

For each skin lesion three elements are necessary. The first one is the patients' consent form. This means the patients let us use the images of all of their skin lesions for teaching purposes. The second element is the image of a skin lesion and the third one is the dermatologist's judgment about the lesion beside the dermoscopy picture. The dermatologists can take pictures of each element using their smart phone. Later, using the application, they can select these images from the gallery and send them to the database. For each patient, the dermatologist can send many images to the server using the same consent form.

This application has been developed for both Android and iOSand both of them have the scheme of Figure 3.2.

3.1.1 Android Data Collector

Before I started working on this thesis, an Android application had been written by Farrukh Ahmed that does exactly what we need in this step. There is

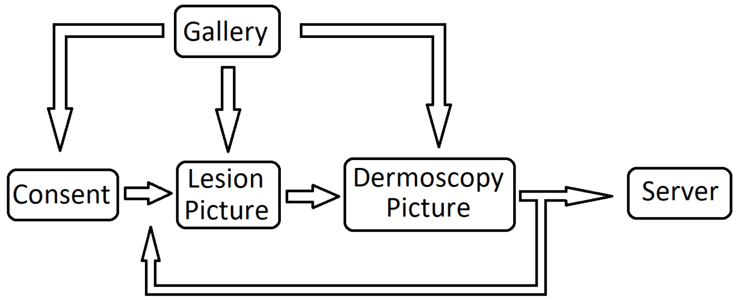


Figure 3.2: The data collector design

a full explanation of his application in the following.

The snapshot below is the homepage of the data collector application. When the dermatologists open the application they need to select a few images and send them to our server. The first image is the consent form of the patient. We need this because the images of the database have to be legal.

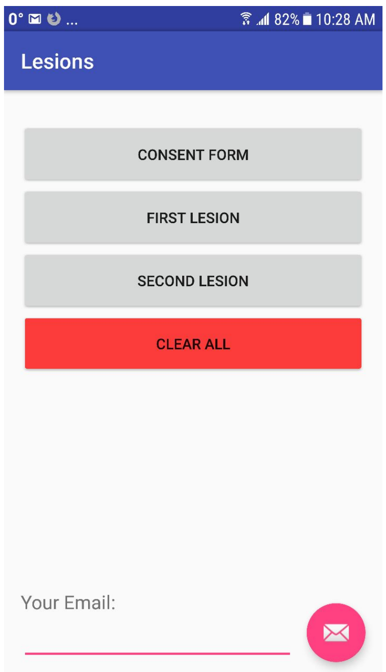


Figure 3.3: The homepage of Android data collector

When the dermatologist clicks on consent form a page will show up which lets them chose an image from their gallery. Below, you can see before and after selecting the image in Figure 3.4.

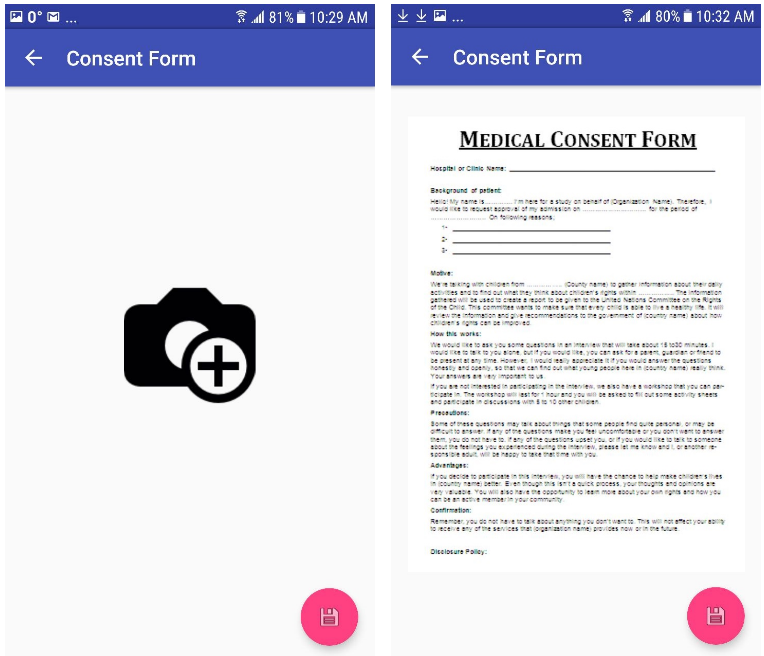


Figure 3.4: The consent form page of Android data collector

After saving the image by clicking on the red button, the user will go back to the homepage where he can now select the images related to the lesion. When he clicks on FIRST LESION button, another page will show up which lets him select three different images. One for the lesion itself, one for the dermoscopy picture and one for a form including the dermatologist's opinion about the lesion. The last part is what we use for putting the image in one of the categories in the database. Below you can see the lesion page before and after choosing an image.

As you can see in the images, there are two more pages for lesion form and for dermoscopy picture which are identical to this page. Once all of the images have been selected and saved, the user goes back to the homepage. He has the option of sending one more lesion to the server with the same consent form. This can be helpful when a patient has two skin lesions.

The only piece left is the identification. It is good to know who has sent the images to the server. If we do not know the dermatologist, we might use images from unknown sources which might not be correctly labeled all the time. There is a bar at the bottom of the page where the user can enter his

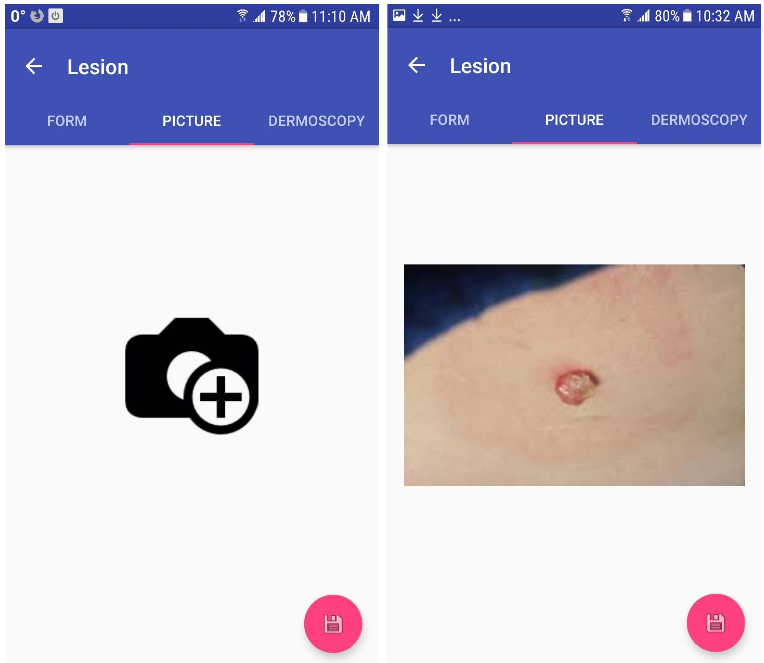


Figure 3.5: The lesion selection page of Android data collector

email address. This is how it looks like when the user is ready to send the lesion(s) to the server.

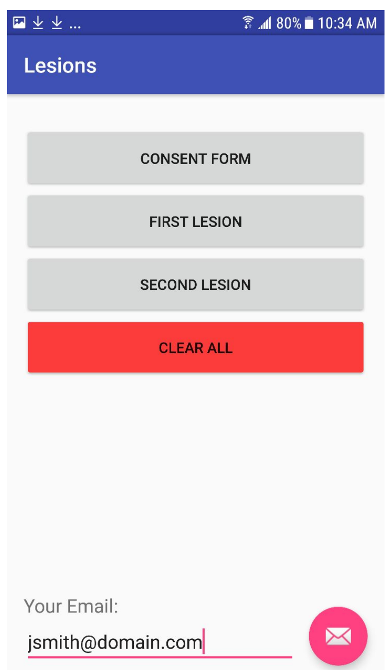


Figure 3.6: The complete homepage of Android data collector

When the images have been chosen and the email address has been filled, the user can send them to the server by clicking the red button at the lower right of the homepage. Then the images will be saved in the database and waiting for a confirmation by an intern who removes lesions with unclear forms or images and approves the useful ones.

This application can be downloaded from our server's webpage using a password.

3.1.2 iOS Data Collector

For writing the iOS application we decided to imitate the design of the Android application so if any of the dermatologists changed their phone from one to another they are still able to use the application easily. So the functionality of the iOS application is the same as the Android application. In the following, we will take a look at the interface of the iOS application.

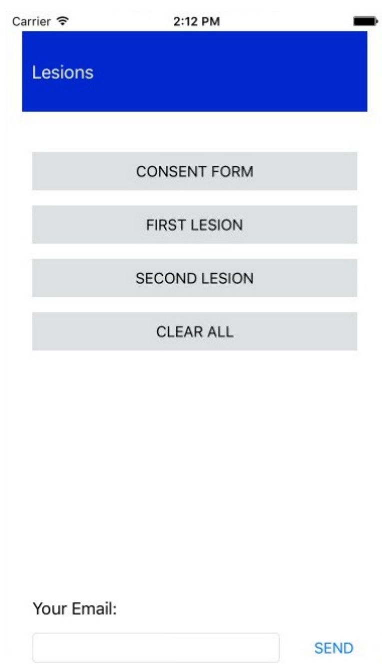


Figure 3.7: iOS Data Collector's homepage

There are not many differences between the iOS application and the Android application, but in the following, we will review some minor differences.

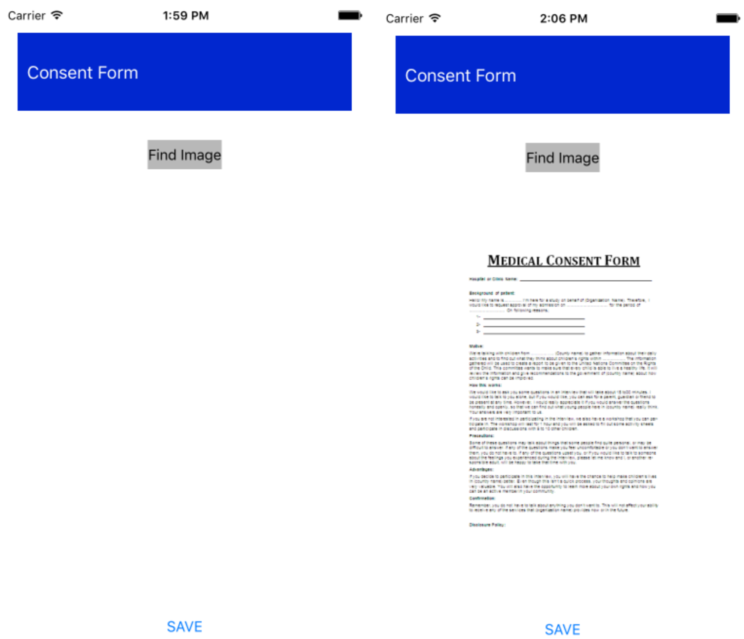


Figure 3.8: iOS Data Collector's consent form page. Before and after choosing the image

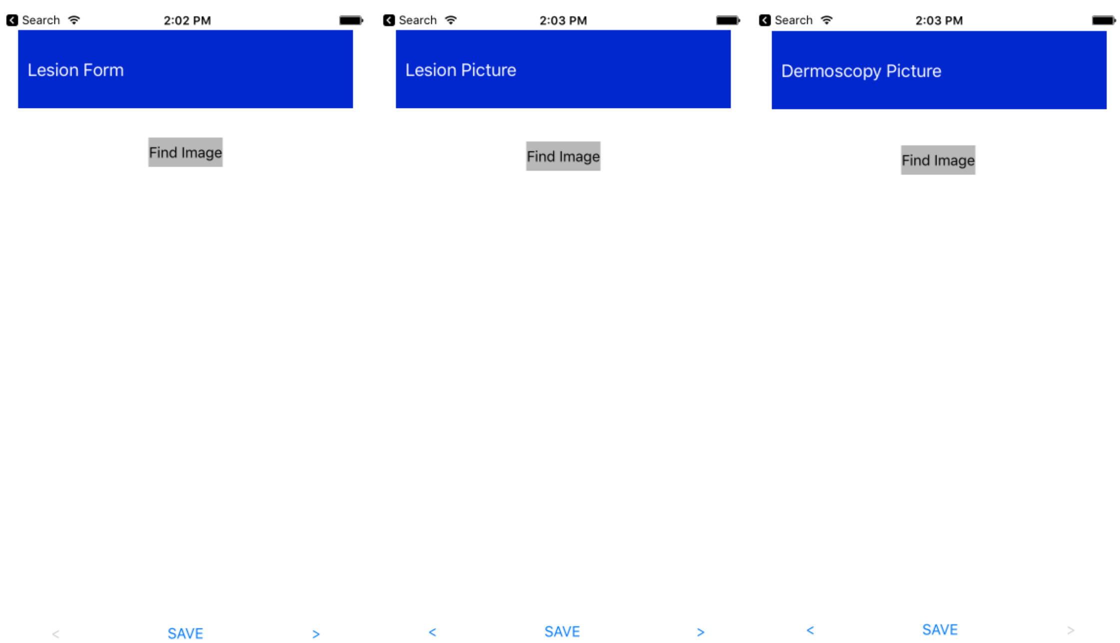


Figure 3.9: iOS Data Collector's lesion selection pages

Since almost everybody has their own smartphone, they will probably use the same phone every day. So it would be easier if they did not have to enter

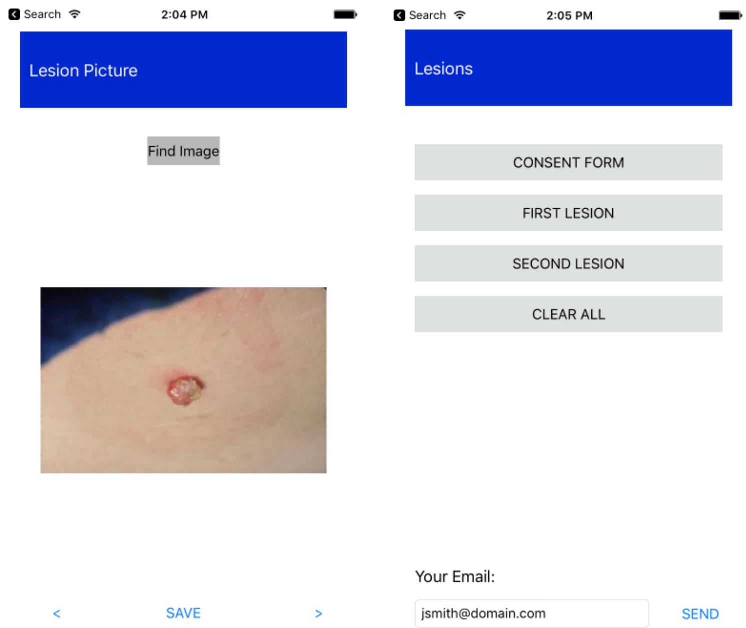


Figure 3.10: a) Lesion page after selectiong the lesion, b) The completed homepage ready for sending the lesions to the server.

the same email address on the same device over and over again. Thus the iOS application remembers the last email address you entered and it does not want you to enter it every time unless you want to change it yourself.

The most important difference between the two applications is that Android applications can be moved from one device to another but iOS applications can not. The only way to get an iOS application is to download it from App Store. It means everybody has access to the iOS application. Thus for security purposes another page was added at the beginning of the iOS application (before homepage) which asks for a password. Once you enter it correctly, it will not be asked the next times.

3.2 Pretest/ Post-test

Now that we have the images in the database, It is time to think about the learning, but before that, We will explain another application. In our experiment which will be explained completely later, we want to measure the improvement of participants. In order to be able to do so, we need an application which can give us a number on how good a participant is in distinguishing skin lesions.

Before explaining the application, it might be useful to remember the categories of skin lesions. There are two main groups of skin lesions: benign and malignant. Each of them has its own subcategories. The subcategories of benign are Lentigo, Blue Nevi, Seb Ker and Acquired Melanocytic Nevi and the subcategories of malignant are Lentigo Maligna Melanoma, Acral Lentiginous, Nodular Melanoma and Superficial Spreading Melanoma.

The pretest application is a simple test on skin lesions in basic level which includes images from all 4 subcategories of both benign and malignant. A sample of the application exists below in Figure 3.11.

Figure 3.11: The pretest and post-test application

This application has two different purposes. The first one is that we need to make sure that our participants do not know much about skin lesions before they start learning. Otherwise, we can not compare their knowledge after the experiment, because their performance can be affected by their knowledge before they started to experiment. So we expect them to have an accuracy close to random. If they do not, they will not qualify for the experiment.

The second purpose of the experiment is measuring their performance before and after the experiment. If we want to compare basic level learning and subordinate level learning we need to compare the participants' improvement since the beginning of the experiment. This application computes the participant's accuracy when the test is finished. The test has 24 images which include 3 images of each subtype of benign and malignant lesions. This assures the test covers all of the subtypes so the participant's performance in this test can show how good they are in classifying skin lesions.

Now we have the tool for selecting useful participants for our experiment. The next step is the teaching application itself which will be explained in the next section.

3.3 Teaching Application

So far we have the tool for collecting skin lesion images. Assuming that the dermatologists will use the application and send lesion images to the database, we can use those images for teaching medical students.

Keep in mind that our question is whether the basic level learning is better than subordinate level learning or is it the reverse? In the basic level learning you learn if an image is benign or malignant, but in the subordinate level learning, you learn the subtype of an image (e.g. this lesion is Seb Ker). In order to answer our question, we need to have a tool which teaches both in the basic level and in the subordinate level.

In the previous section, we talked about the subtypes of benign and malignant skin lesions, but you should also know each of the subcategories of benign pairs with one of the subcategories of malignant. For example, Lentigo pairs with Lentigo Maligna Melanoma. This means that the distinction between these two are not really easy while the distinction between Lentigo and Nodular Melanoma might not be difficult.

Since learning all of the categories in a single session seems to be hard and confusing it was decided to have multiple sessions. In the basic level, the user sees 16 pairs of images. Each pair consists of a benign and a malignant lesion. In each session, the user will be shown just one subtype of benign versus the subtype of malignant which pairs with that, but the labels are in basic level (benign or malignant). On the other hand, in the subordinate level, the user learns one subtype of either benign or malignant (such as Lentigo) by seeing an image of that subtype versus an image of another subtype of the same type. For example, in the first pair, Lentigo is shown versus Blue Nevi and in the second pair it is shown versus Seb Ker. This lets them find the differences between Lentigo and other subtypes.

There are 4 subtypes of benign and malignant, Although, In this application, only three of them are used. Because another factor that can be measured at the end of the experiment, is their ability on expanding their knowledge to the new types of lesions. This is important because if they are able to expand their knowledge to a new type of lesion they are more likely to correctly classify unseen images from the same types.

Now we know two features that our application needs to have. The first one is that it has to teach both in the basic level and in the subordinate level, and the second one is that it should have 3 sessions.

Since the development of this system has not been only for the experiment and since we want to be able to use this system for teaching medical students, one part of the system is designed for the second goal, i.e. teaching. Assuming that we have more than enough skin lesion images for teaching medical students, we can create more than one session for each subtype, but if we decide to do so, it is better to have the easier images in the first session and the harder images in the next sessions. In order to do this, we need the difficulty level of each image. For this reason after each session in the teaching application, an assessment has been designed which is on unseen images of the same subtypes of the teaching session. The difficulty level of each image is based on how often it is classified correctly.

There is only a tiny difference between the tests for the basic and the subordinate level. In the basic level the test shows some images and for each of them, it asks if the image is benign or malignant, but in the subordinate level an image will be shown first and then another one after it. The question is if these two images belong to the same subtype. So instead of answering a question for each image, participants of the subordinate level answer a question for each pair. These data are also kept on the server and shows how similar tow images are.

Other than the test after each session, there is another type of test. After each pair has been shown in either the basic level or the subordinate level, there is a short test which asks about the same two images. The user has to select the correct label of the image(e.g. benign in the basic level and Seb Ker in the subordinate level). They can see the feedback after they answer each question. Green shows they have answered the question correctly and red shows they have answered it wrong. If they choose the wrong label the same question will be asked after the next pair. It will keep being asked until the user answers it correctly. This is the only place they get feedback after a question. They get feedback neither in pretest nor in the tests after each session.

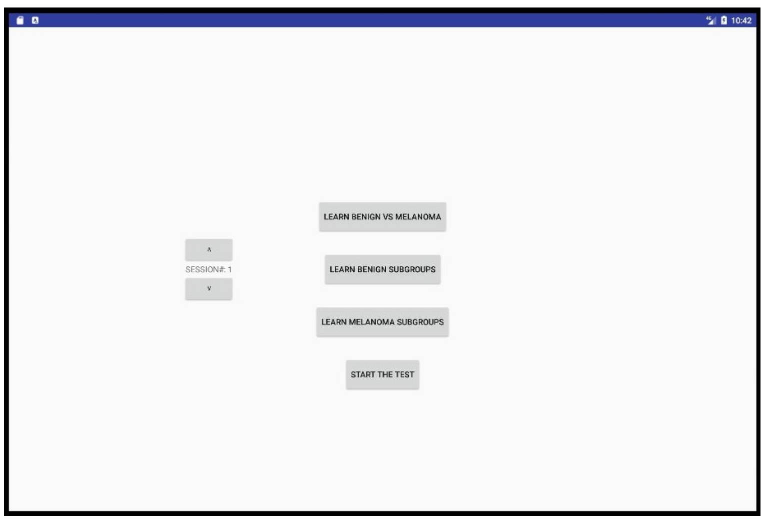
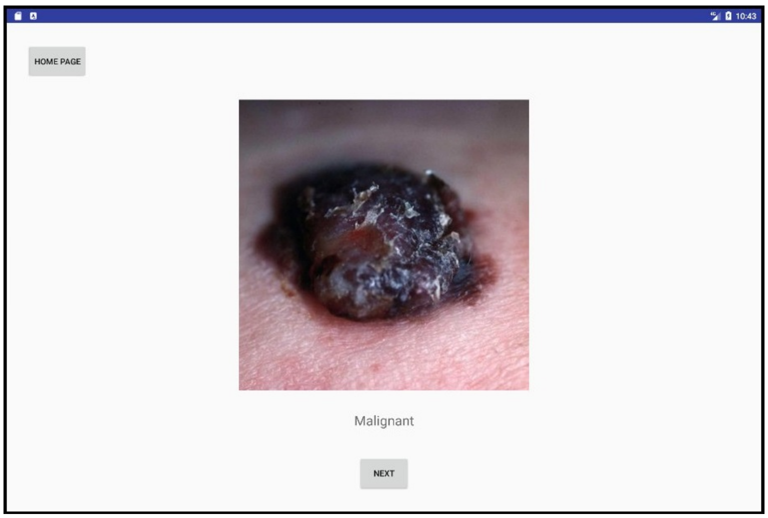


Figure 3.12: The teaching application homepage

3.4 Server

The Server has been implemented using Django, python3 and apache. A part of the server which deals with the Android data collector had been implemented by Farrukh Ahmed. The rest of the functionalities have been added to that server later. As mentioned earlier, the server can get connected to the Android data collector and the iOS data collector. These applications

(a) Teaching Application, the first image of the basic level.



(b) The second image of basic level

Figure 3.13: Basic level teaching

convert the images to strings using Base64 encoder. The string will be sent to the server through HTTP connection. The server then decodes the strings and gets the image and saves it in the memory with the email address of the sender.

The teaching application and the pretest application need to get the image instead of sending it. Thus they need to define the parameters of the image they want. These parameters are the category of the image (benign or malignant), the subtype of the image(e.g. Nodular Melanoma) and the number of the image in that subcategory. The server application queries the database with these parameters and finds the image. Then it uses the same method as was used in the data collectors. It converts the image to a string using Base64 encoder and sends the string to the Android device using HTTP connection. The Android device receives the string and decodes that into an image using Base64 and shows it to the user.

Besides this information, some more data is saved for each image and that is the hardness of the image. That is the number of times it has appeared in the tests in the basic level and in the subordinate level. This also includes the number of times the image has been answered correctly in each of the levels. Therefore, an image can be difficult in the basic level and not so hard in the subordinate level. As mentioned before this information can be used in order to have different sessions on the same subtypes with increasing difficulty.

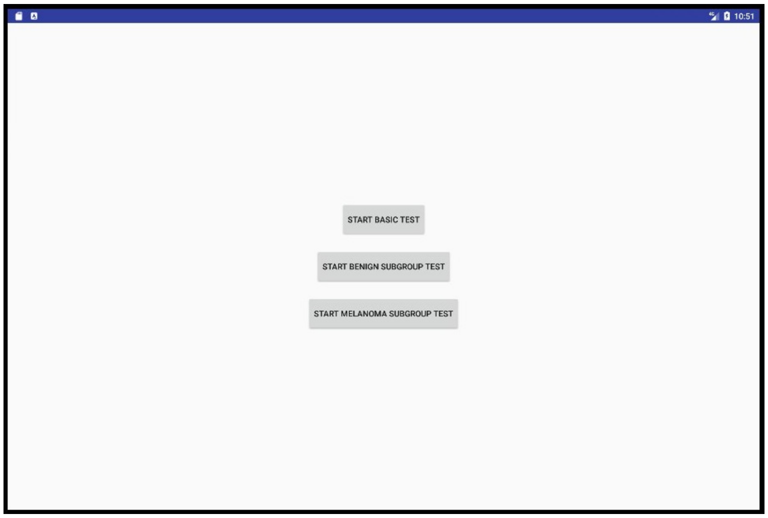
This part of the thesis can be called the most important part of the whole system. Without the server, none of the other applications would work properly. All of them need to communicate with the server and this makes the server significantly important.

Now that all of the parts of the system have been explained we can go to the experiment. Using this experiment, we are trying to answer the question about the best level of learning in skin lesions. The details of the experiment will be discussed more in the next chapter.

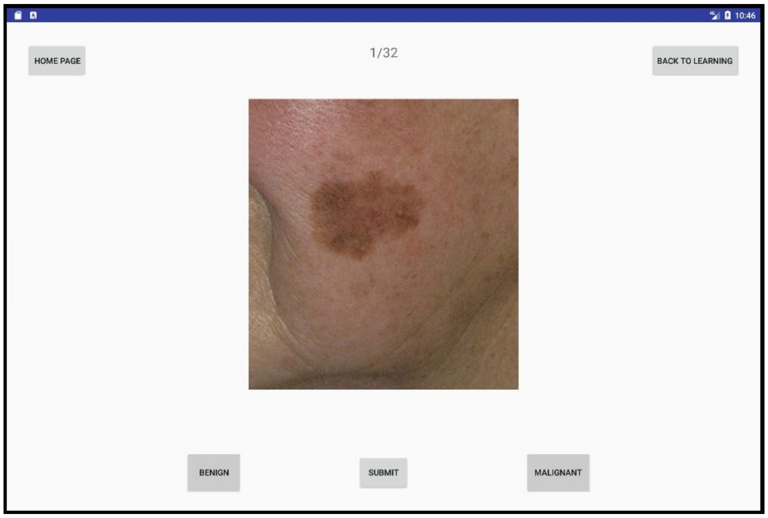
(a) The question after a pair. User has selected benign but he can change his opinion before submitting.

(b) The feedback after submitting a question in basic level.

Figure 3.14: Basic level teaching - questions



(a) The page for selecting the test after a session.



(b) The basic test after a session. This is similar to the questions in teaching session except that the user does not get feedback in this test.

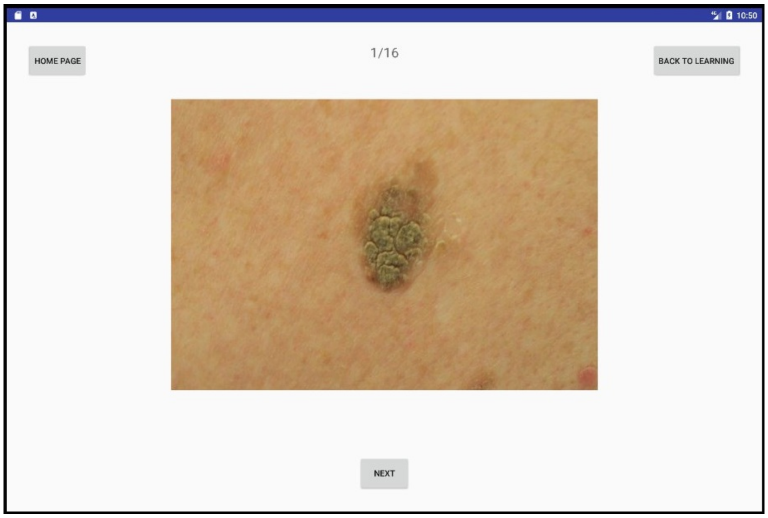
Figure 3.15: Basic level test

(a) The teaching session of subordinate level - part 1

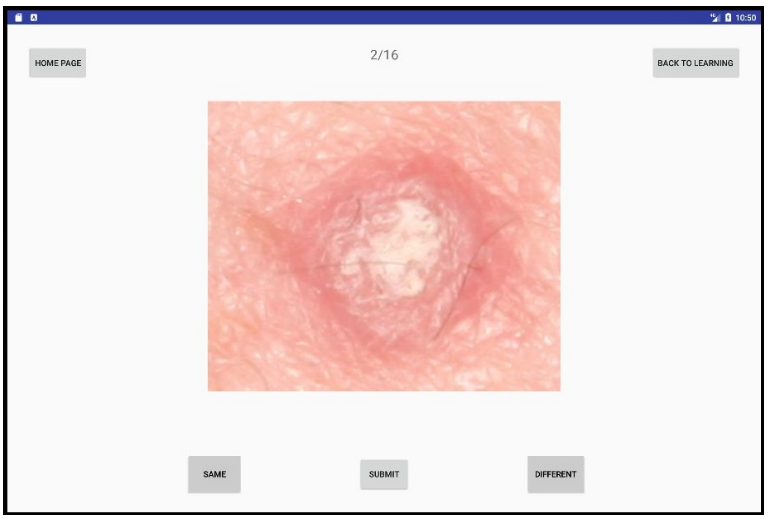
(b) The teaching session of subordinate level - part 2

(c) The question after each pair in subordinate level learning.

Figure 3.16: Subordinate level teaching



(a) The subordinate test after a session. The first image in the pair.



(b) The subordinate test after a session. The second image in the pair. The user has to say if this image is in the same subtype as the previous image. There is no feedback in this part.

Figure 3.17: Subordinate level test

Chapter 4

Experiment

As stated earlier, the question I am trying to answer is about the best way of learning skin lesions. Which one of the basic level and the subordinate level learning is more effective? In order to answer this question, an experiment has been designed. In this experiment, there are two groups of participants. One of them learns the skin lesions in the basic level and the other group learns in the subordinate level. Then we compare their performance at the end at the basic level to see which one of them has learned better; i.e. performs better on post-test. In the following, I will talk about each part of the experiment separately.

4.1 Pretest

Pretest needs to be performed to make sure that none of the participants knows about distinguishing between melanoma and normal cases. I do not want the participants to know distinguishing lesions because if they do, the comparison which I will have at the end, will not be meaningful. Besides, I need to have the participants' baseline abilities before training, so that we can estimate the effect that training has on their abilities.

4.1.1 Participants

The number of participants in this experiment is very low. There were only five participants in this experiment, which may be due to the fact that this experiment is done in the course of four days for each participant. The results of the experiment would be more reliable if we had more participants, but having few participants let us have a rich data about each one of them.

4.1.2 The Lesions in the Pretest

The pretest consists of 24 different images. For each of the subtypes, we have three images in the pretest. That means there are 12 benign images and 12 malignant images in the pretest.

4.1.3 Evaluation Metrics

For measuring the performance of the participants we use 5 different metrics. These metrics are as follows.

- **Accuracy:** This is the most common metric for measuring the performance. The formula of accuracy is: $A = \frac{TP+TN}{TP+FP+TN+FN}$
- **Sensitivity Index:** This metric measures the participant's performance compared to others: $d' = \frac{x-\mu_N}{\sigma}$
- **Precision:** This shows the participant's performance among only the lesions that he answered as malignant: $p = \frac{TP}{TP+FP}$
- **Recall (Sensitivity):** This shows the participant's performance among all malignant lesions: $r = \frac{TP}{TP+FN}$
- **Specificity:** This shows the participant's performance among all of the benign lesions: $s = \frac{TN}{TN+FP}$

In these metrics, μ_N shows the mean, σ is the standard deviation, TP (True Positive) is the number of malignant lesions which the participant answered correctly, TN (True Negative) is the number of benign lesions that the participant answered correctly, FP (False Positive) is the number of benign lesions that the participant answered incorrectly and FN (False Negative) is the number of malignant lesions which the participant answered incorrectly.

4.1.4 Pretest Performance

The pretest performances of the participants were 58% (d'=-0.328939248, p= 0.57, r= 0.67, s= 0.5), 58% (d'=-0.745233063, p=0.58, r=0.58, s= 0.58), 58% (d'=-0.328939248, p=0.56, r=0.75,s= 0.42), 62% (d'=1.40311156, p=0.6, r=0.75, s=0.5) and 83%(p=0.79, r=0.92, s= 0.75). Obviously, we did not continue the experiment with the participant with 83% accuracy (Which is why there is no d' score for this participant). The question that can be asked here is why all of the participants had an accuracy higher than 50%. The answer is that there are some lesions that obviously have a problem. You do not know what the problem is, but once you look at the lesion, you know something is wrong. An example of such lesions exists below in Figure 4.1.

Figure 4.1: An example of lesions which are obviously malignant.

You do not need to be an expert to know that this lesion is malignant. Considering that there are few images like this in the pretest, it seems only reasonable for the participants to have an accuracy slightly higher than 50%. For these obvious images they all select malignant and for the other images, their answer seems to be randomly selected. Not only do these numbers show that the participants do not know much about the lesions, but also show that they have done their best and have answered the obvious images correctly.

4.2 Learning

After doing the pretest, There were four participants left for continuing the experiment. In this step, the participants are divided into two groups. One of them is learning the skin lesions in the basic level and the other learns them in the subordinate level.

In the basic-level condition, in each session, one of four types of benign lesions and one of the four types of malignant lesions are used to teach users to distinguish between benign and malignant lesions. Then they are tested on this distinction, and after a 24-hour delay, they begin a new round of training using a second type of benign lesions and a second type of malignant lesions.

In the subordinate level, the user learns one subtype of benign (malignant) versus other subtypes of benign (malignant). Remember that we are keeping one subtype of benign and one subtype of malignant for the post-test, so when we say other subtypes of benign, it means only two other subtypes. After the learning session in the subordinate level, the user does the test for the first session. Then he waits for 24 hours and then goes to the next session.

The tests after each session are not necessary for the learning process, because the users will not get a feedback in them. These tests are useful because they help us gather some information about the images such as the difficulty of each image. This data can help us form the sessions in future. That means if we have enough images for even more sessions, the easier images will be in the first session and the harder ones will be in the second.

All three sessions of the experiment have been done with the four participants. Two of them learned the skin lesions at the basic level and two of them learned them at the subordinate level. Before looking at the results for the post-test, we will discuss an observation on their performance in the tests at the end of each session.

Both of the participants in subordinate level had an acceptable performance in the session which taught them malignant subtype Nodular Melanoma versus the other two (75% and 100%), but both of them had a lower performance in the sessions for Acral Lentiginous (37% and 62%) and Lentigo Maligna

Melanoma (50% and 50%). This suggests that Acral Lentiginous and Lentigo Maligna Melanoma are really similar and recognizing the difference between them is not an easy job.

4.3 Post-test and Results

All of the participants have had a better performance in the post-test compared to the pretest. The participants of the subordinate level had the accuracy of 71% (d'=-2.027920408, p=0.69, r=0.75, s= 0.67) and 75% (d'=-1.253323739, p=0.71, r=0.83, s= 0.67) and the participants of the basic level had the accuracy of 92% (d'=1.253323739, p=0.92, r=0.92, s= 0.92) and 96% (d'=2.027920408, p=0.92, r=1, s= 0.92). This means that on average, the participants of the basic level had 34% of improvement in their performance while the participants of the basic level only had only 15% of improvement on average. The number of the participants is not enough for a reliable conclusion but there is an observation here:

The participants of the basic level learned skin lesions better than the participants in the subordinate level. This is exactly in contrast with the study of birds in which the subordinate level learning showed to have a better effect on the participants.

That is not really strange, because the category structure of birds and skin lesions are different. When you look at the phylogenetic tree of birds (Figure 4.2) , you see that all of the different subtypes of owls have been created based on the first type of owls and all of the different subtypes of wading birds have been created based on the first type of wading birds. You can also see that owls and wading birds are siblings in the higher level. This means when you study wading birds at the subordinate level you are learning the full subtree of wading birds in the birds' phylogenetic tree.

On the other hand, many benign lesions are more similar to malignant lesions than they are to benign lesions from the same subtype. Thus based on these lesions, one can say that skin lesions are divided into four types of lesions first, then each of the subtypes is divided into two groups of benign

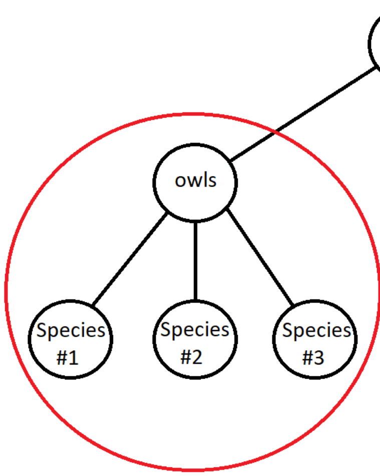


Figure 4.2: Birds phylogenetic tree

and malignant (Figure 4.3). This means that in the phylogenetic tree of skin lesions, Lentigo is closer to Lentigo Maligna Melanoma compared to Blue Nevi which is exactly the opposite of the case in the birds' category structure.

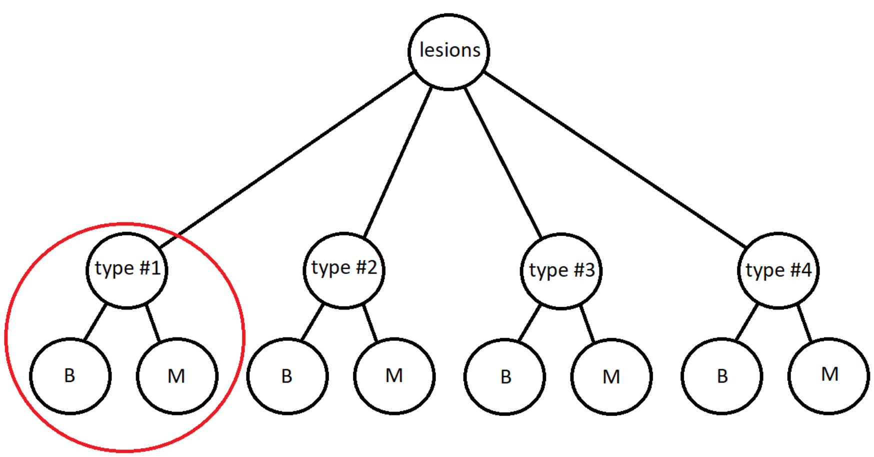


Figure 4.3: Lesions phylogenetic tree

This also suggests that the category structure of skin lesions is dispersed because the members of benign, are sometimes more similar to members of malignant. This is also compatible with the subordinate level learning being better than the basic level learning.

There is another exciting observation in this experiment. If you pay attention to the performance of the participants in the basic level you see that out of 24 images in the post-test, one of them has answered 22 and the other has answered 23 images correctly. Considering the images in the post-test, this means that out of the 6 images in the fourth subtype of benign and the fourth subtype of malignant, they have each answered 5 images correctly. This fact suggests that they have been able to generalize their knowledge for the subtype that they never learned. This is really interesting because it means that they might be able to correctly identify the skin lesions that they see in the future which is the final goal of this project. In Table **??**, you can see the number of correct answer of each participant in each subtype in both pretest and post-test.

Of course, we need to have a lot more participants in the experiment in order to claim these observations with more certainty. The low number of the participants are only due to the tight timeline and repeating the experiment with more participants is the most important part of the future work.

		Participant #1	Participant #2	Participant #3	Participant #4
Pretest	Lentigo	1	2	1	2
	Lentigo M.M.	3	2	2	2
	Blue Nevi	2	1	2	1
	Acral L.	2	3	2	2
	Seb Ker	1	1	3	2
	Nodular M.	3	2	2	1
	Acquired M.N.	2	1	0	2
	Supercial S.M.	1	2	2	2
	Sum	**15**	**14**	**14**	**14**
Post-test	Lentigo	3	3	3	2
	Lentigo M.M.	3	3	3	3
	Blue Nevi.	3	3	2	2
	Acral. L.	3	3	2	3
	Seb Ker	3	2	2	3
	Nodular M.	3	3	3	2
	Acquired M.N.	2	3	1	1
	Supercial S.M.	3	2	1	2
	Sum	**23**	**22**	**17**	**18**

Table 4.1: Detailed results

Chapter 5

Conclusion and Future work

In this thesis we reviewed a few ways to learn skin lesions. Then we read about two experiments on the best level of learning in different subjects. The family-resemblance principle was introduced as well as compact and dispersed category structure. In the third chapter the tools for a new experiment were explained and in the fourth chapter we went over the experiment.

There are some observations based on the results of the experiment. The first one is that in the malignant lesions, Acral Lentiginous and Lentigo Maligna Melanoma are really similar and Nodular Melanoma is fairly distinguishable among them. The second observation is that the knowledge on the categories of skin lesions (the basic level) can be generalized to unseen images in the same categories and unseen images in new categories. In order to verify this statement similar experiments should be done with different subtypes of lesions as the new category. It means that there need to be three more experiments similar to the one which was done in this thesis.

The most important observation is that unlike the results of the study on birds [11], in skin lesions, learning in the basic level is more efficient that learning in the subordinate level. We also saw that the skin lesions have a dispersed category structure which is compatible with the results of the study on rocks [7]. This observation is based on an experiment on only 5 participants. Thus the next step would be doing the same experiment but with many more participants. The result would be much more reliable in that case.

There are seven different approaches to learn skin lesions: the dermatology elective, multicomponent interventions, computer-based learning, lecture, pamphlet, audit and feedback and moulage. The computer-based methods are ranked as the third most effective method for learning the skin lesions [9]. One can compare the effectiveness of the proposed method in this thesis to other approaches, to see if this method can improve the computer-based methods' rank in learning skin lesions.

One of the obstacles in this project was dermatologists not using the data collector applications. The application might seem time consuming for a dermatologists. One of the possible solutions for this problem can be to simplify the process for them, like adding the option of taking images from within the application or grouping the images by category. It makes the images easier to be found since they are not in the gallery among many other images which are usually not related to skin lesions. This can reduce the time needed for sending a lesion to the server as well as the error rate in selecting the images.

In order to use the whole system the data converter need to be implemented. This component should be provided to an intern to verify the images coming to the server and remove the lesion images which have a problem. In this thesis an existing database was used. Without this component the data collector applications are not useful.

Bibliography

[1] Naheed R Abbasi, Helen M Shaw, Darrell S Rigel, Robert J Friedman, William H McCarthy, Iman Osman, Alfred W Kopf, and David Polsky. Early diagnosis of cutaneous melanoma: revisiting the abcd criteria. *Jama*, 292(22):2771–2776, 2004.

[2] Linda Argote, Paul Ingram, John M Levine, and Richard L Moreland. Knowledge transfer in organizations: Learning from the experience of others. *Organizational behavior and human decision processes*, 82(1):1–8, 2000.

[3] Shlomo Bentin and David Carmel. Accounts for the n170 face-effect: a reply to rossion, curran, & gauthier. *Cognition*, 85(2):197–202, 2002.

[4] Robert J Friedman, Darrell S Rigel, and Alfred W Kopf. Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. *CA: a cancer journal for clinicians*, 35(3):130–151, 1985.

[5] Isabel Gauthier, Pepper Williams, Michael J Tarr, and James Tanaka. Training greebleexperts: a framework for studying expert object recognition processes. *Vision research*, 38(15):2401–2428, 1998.

[6] Mary E Lassaline, Edward J Wisniewski, and Douglas L Medin. 9 basic levels in artificial and natural categories: Are all basic levels created equal? *Advances in psychology*, 93:327–378, 1992.

[7] Robert M Nosofsky, Craig A Sanders, Alex Gerdom, Bruce J Douglas, and Mark A McDaniel. On learning natural-science categories that violate the family-resemblance principle. *Psychological science*, 28(1):104–114, 2017.

[8] Jennifer J Richler and Thomas J Palmeri. Visual category learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(1):75–94, 2014.

[9] Liam Rourke, Sarah Oberholtzer, Trish Chatterley, and Alain Brassard. Learning to detect, categorize, and identify skin lesions: a meta-analysis. *JAMA dermatology*, 151(3):293–301, 2015.

[10] Stefan R Schweinberger, Vyv Huddy, and A Mike Burton. N250r: a face-selective brain response to stimulus repetitions. *Neuroreport*, 15(9):1501–1505, 2004.

[11] Lisa S Scott, James W Tanaka, David L Sheinberg, and Tim Curran. A reevaluation of the electrophysiological correlates of expert object processing. *Journal of cognitive neuroscience*, 18(9):1453–1465, 2006.

[12] Robert S Stern. Prevalence of a history of skin cancer in 2007: results of an incidence-based model. *Archives of dermatology*, 146(3):279–282, 2010.