

AMERICAN UNIVERSITY OF BEIRUT

mHEALTH SYSTEM FOR DERMATOLOGY
DISEASES IN REFUGEE SETTLEMENTS
USING MULTI-MODAL CLASSIFICATION

by

FADY GEORGES BALY

A thesis

submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
to the Department of Electrical and Computer Engineering
of the Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon

September 2020

AMERICAN UNIVERSITY OF BEIRUT

mHEALTH SYSTEM FOR DERMATOLOGY
DISEASES IN REFUGEE SETTLEMENTS
USING MULTI-MODAL CLASSIFICATION

by

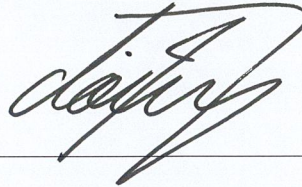
FADY GEORGES BALY

Approved by:

Dr. Zaher Dawy, Professor

Advisor

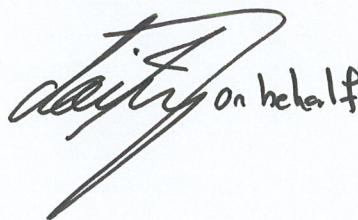
Electrical and Computer Engineering



Dr. Hazem Hajj, Associate Professor

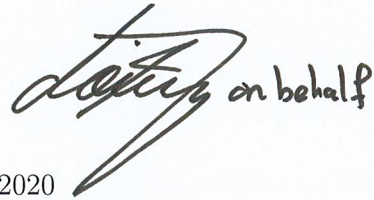
Member of Committee

Electrical and Computer Engineering



Dr. Mazen Kurban, Associate Professor
Biochemistry and Molecular Genetics

Member of Committee

 on behalf

Date of thesis defense: September, 4, 2020

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT

RELEASE FORM

Student Name: Baly Fady Georges
Last First Middle


Master's Thesis Master's Project Doctoral Dissertation

I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes after: **One** ___ year from the date of submission of my thesis, dissertation or project.

Two ___ years from the date of submission of my thesis, dissertation or project.

Three years from the date of submission of my thesis, dissertation or project.



Signature

Date

Sept. / 17 / 2020

This form is signed when submitting the thesis, dissertation, or project to the University Libraries

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Prof. Zaher Dawy for his motivation, knowledge, and continuous support, which was a key ingredient towards successfully completing my Thesis.

I have also been fortunate to work under the supervision of outstanding scholars: Prof. Hazem Hajj (AUB), Prof. Mazen Kurban (AUB) who were generous to provide me with guidance and insightful comments. I am also thankful to all the members in the AUB Mind Lab who helped a lot through collaborations and discussions.

Finally, and most importantly, I would like to thank my brother Ramy for his support, patience and tolerance, especially in patches of rough times. I would also like to thank Ruby, she's been a strong pillar as she constantly motivated me to finish this chapter in my life through her words of encouragement and motivating speeches. I also want thank my parents, Georges and Caroline, for their moral support and their faith in me and allowing me to be as ambitious as I wanted.

An Abstract of the Thesis of

Fady Georges Baly for Master of Engineering
Major: Electrical and Computer Engineering

Title: mHealth System for Dermatology Diseases in Refugee Settlements Using Multimodal Classification

While conflicts and wars are continuously erupting throughout the world, the dispersion of refugees from their war-afflicted countries to neighboring states is continuously increasing to an extent that these hosting states become incapable of meeting the refugees' basic needs, such as shelter, water, education, and most importantly, healthcare. In particular, the hardships that refugees have been facing due to the lack of basic medical services have motivated researchers to develop technological automated solutions to address existing challenges and enhance healthcare support.

The focus of this thesis is on developing a mobile health system for diagnosing certain skin diseases in an automated and accurate manner. The proposed system leverages recent advances in machine learning, in particular deep learning algorithms that are applied to different data modalities. The system's architecture includes a user question-answer component to retrieve user-related background and health information, followed by an embedding model that is used to learn representations for uploaded images capturing the affected areas of their skin.

Finally, a machine learning classifier is trained using the features extracted from both the questionnaire and image modalities to accurately predict the type of skin disease, providing preliminary input to remote medical experts for further evaluation and treatment. Testing and evaluation are performed using various real data sets, and the obtained results demonstrate the overall effectiveness of the proposed approach.

Contents

Acknowledgements	v
Abstract	vi
1 Introduction	1
2 Literature Review	5
3 System Description	9
3.1 System Architecture	9
3.2 System Modules	10
3.2.1 Image-based Diagnosis	11
3.2.2 Form-based Diagnosis	12
3.2.3 Expert Diagnosis	12
3.2.4 Mass Diagnosis	13
4 Dataset and Models	14
4.1 Datasets	14
4.1.1 SYNFORM Data	14
4.1.2 DERMNET Data	21
4.2 Data Transformation	22
4.2.1 Images: DERMNET	22
4.2.2 Form: SYNFORM	23

4.3	Image Classification Models	24
5	Experiments and Results	28
5.1	Experimental Setup	29
5.1.1	Baseline	29
5.1.2	CNN Standalone	30
5.1.3	Majority Voting	31
5.1.4	Fusion: CNN with Form Features	32
5.2	Evaluation Metrics	32
5.3	Experimental Results	34
5.4	Results Analysis	36
6	Conclusion and Future Work	41
A	Smartphone Sensors for mHealth	44
B	Machine Learning Models	47
B.1	Support Vector Machines (SVM)	47
B.2	Convolutional Neural Networks (CNN)	48
B.3	Entity Embedding	50

List of Figures

3.1	System architecture of our proposed MHEALTH method.	10
3.2	The image-based diagnosis module.	11
3.3	The form-based diagnosis module.	12
4.1	The distribution of the different classes in the SYNFORM dataset.	17
4.2	The distribution of the different classes in the DERMNET dataset.	22
5.1	Majority Voting approach structure.	31
5.2	Feature Fusion approach structure.	32
5.3	Comparison between cases with a) mild and b) visible symptoms [1].	38
5.4	Comparison between a symptoms on white skin (a), and symptoms on dark skin (b) [1].	39
5.5	Example of an image with large covered area of the body [1].	40
B.1	The maximal marginal hyperplane in SVM. source:	48
B.2	The architecture of a CNN including the different types of layers.	49
B.3	The convolution operation. source:	49
B.4	The pooling operation. source:	50
B.5	Entity Embedding operation [2]	51

List of Tables

4.1	Summary of the different features that exist in the synthetic dataset.	
	15	
4.2	CNN models specifications.	27
5.1	Results of the baseline model that only uses the “Form” features.	35
5.2	The individual performance for the different CNN classifiers. . . .	35
5.3	The performance of the ensemble CNN model with and without incorporating the Form-based model.	36
5.4	The performance of the EfficientNet-B7 (Noisy Student) model with and without the Form features.	37
5.5	F_1score per class comparison between SVM and CNN ensemble .	38
5.6	F_1score per class EfficientNet and EfficientNet with form features	40
A.1	Overview of the capabilities of the different smartphone sensors for mHealth applications [3]	46

Chapter 1

Introduction

The flow of refugees from their war-afflicted countries to neighboring states is continuously dispersing. For example, according to the United Nations High Commissioner for Refugees (UNHCR), the number of Syrian refugees in Lebanon has surpassed the 1 million mark,¹ which is significant compared to Lebanon's population of 4 million. Such situations would eventually result in the inability of the hosting states to provide refugees with basic needs, including shelter, food, education, and medical care. In particular, the scarcity of healthcare services would create further hardships for refugees in need of medical care, especially those who reside in primitive camps that are not equipped with the necessary medical infrastructure to provide basic medical services such as diagnosis and treatment. In addition to the lack of medical resources and attention, the language barrier between refugees and physicians, if applicable, is another serious challenge that might hinder proper communication, which in turn would lead to inaccurate diagnosis and treatment [4].

Dermatological (or skin-related) diseases are commonly widespread in refugees camps. Studies presented at the 2016 European Congress of Clinical Microbiology and Infectious Diseases (ECCMID) revealed a list of dermatological diseases that

¹<http://data.unhcr.org/syrianrefugees/country.php?id=122>

are widespread amongst refugees in the Netherlands and Switzerland.² Those observed diseases represented a burden to the European States' healthcare systems, as they require extensive monitoring of refugees, especially those coming from Africa and the Middle East. Furthermore, those studies found that 30% of African refugees have shown evidence of scabies. Another study³ have documented the spread of scabies among a large number of Syrian refugees in Lebanon. Since many dermatological diseases are contagious, early diagnosis and treatment are crucial to prevent widespread dissemination. Based on this information, providing refugees access for early diagnosis of dermatological diseases is of utmost importance from a public health perspective.

Smartphones have recently become the most acquired devices, with the number of users constantly spiking in both developed and underdeveloped countries [5]. Therefore, significant efforts have been invested to exploit this technological advancement for the better of humanity. This includes developing mobile applications that can provide a wide range of services to improve the people's quality of life. Most applications are contingent upon the availability of sophisticated built-in hardware, such as sensory systems that can collect, in real-time, personalized data such as location, image, video, and vital signs. In particular, mobile health applications (or mHEALTH apps) are developed with the purpose of providing healthcare and medical care services, such as obesity, depression, and smoking [6]. These apps rely on a variety of portable and wearable network-capable gadgets, in addition to smartphone built-in features such as microphones, cameras, GPS, and Bluetooth. A detailed description of smartphone features used in mHEALTH apps is available in Appendix A. mHEALTH apps have a great impact on public health, and have even a greater impact in underdeveloped countries as they help remedy problems caused by the lack of decent clinical resources and healthcare support [7].

²<https://www.healio.com/infectious-disease/emerging-diseases/news/online/>

³<https://en.annahar.com/article/633416>

The aim of this thesis is to develop a mHEALTH system that provides early and preliminary diagnosis of dermatological diseases and skin lesions. This system would be of great value for refugees who have limited or no access to medical and healthcare services. Also, because dermatological diseases can easily spread in refugee camps, this app can help prevent epidemic outbreaks through early infection detection. Most mHEALTH apps still rely on actual physicians to analyze images that are uploaded by the users (patients). This is considered a problem as analysing images does not provide enough concordance compared to checking up on the patient vis-à-vis [8]. In other words, it might result in loss of information leading to wrong diagnosis. Recently, mHEALTH apps started to exploit and integrate the capabilities of machine learning into their systems, especially after the booming of the deep neural networks. These apps rely only on images captured by the users, without further information that are typically acquired by the physician during the examination, that might be essential for proper diagnosis.

Accordingly, our proposed system is mainly based on machine learning and deep learning algorithms that are applied to data in different modalities including natural language and images. The system is composed of an interactive survey or questionnaire that the user (patient) has to fill and that helps retrieving relevant user-related information, such as personal history and previous health records. The system also allows users to upload images of the infected parts of their skin. Then, a specifically-trained Convolutional Neural Network (CNN) model [9] is used to extract semantic embedded representations of these images. Finally, a Support Vector Machine (SVM) classifier [10] uses the features extracted from both the questionnaire and the images in order to predict the type of the disease. A detailed description of the different machine learning algorithms that are used by the system is included in Appendix B.

The remaining of the thesis is organized as follows. Chapter 2 provides an overview of existing mHEALTH apps and their underlying mechanisms. The system architecture of the proposed mHEALTH app is then presented in Chapter 3.

This is followed by a description of the machine learning algorithms, as well as the datasets that are used to train these algorithms in Chapter 4. Then, Chapter 5 includes details about the experimental setup, the resulting performances with a discussion Section. Finally, we provide conclusive remarks in Chapter 6.

Chapter 2

Literature Review

As we mentioned earlier in the Introduction (Chapter 1), recent advances in the smartphone technology have led the development of a variety of mobile health applications, or MHEALTH apps, that generally aim to help and guide patients dealing with health-related issues, and to provide medical care services such as self-diagnosing particular diseases. In this chapter, we provide an overview of relevant work that has been conducted in the area of MHEALTH apps.

Overall, MHEALTH apps can be categorized according to the services that they provide, and also according to the means by which they provide these services. One category of these apps help facilitating *direct contact* with physicians. For example, BABYLON¹ has a live chat with a medical expert who provide patients with immediate answers to medical questions related to their health issues. It also offers the option to upload images so that physicians can have a look at the affected area and come up with an accurate diagnosis. Web applications can also serve as health solutions. For example, ICLINI² provides the ability to chat with physicians online, and to book appointments if necessary.

More recent MHEALTH apps rely on artificial intelligence (AI) and use automated “chatbot interface” as a means to interact with the patients and collect

¹<https://www.babylonhealth.com/>

²<https://www.icliniq.com/>

personalized information that can help diagnosing the problem at hand. The collected information can be in the form of (i) answers to questionnaires or (ii) natural language, whether it be speech or text. Different types of chatbots have been used, including smart chatbots [11] and rule-based chatbots [12]. Additional information can be collected via different types of sensors or devices such as scales, heart-rate and blood pressure monitors. For example, the SENSELY³ application uses wearable sensors to provide physicians access to their patients' medical information and vital signs, in order to assess their situation without the need for an in-person examination. This app also offers patients the option to chat with a medically-trained chatbot to provide further information that are relevant to resolve their issues. Other apps, such as MEDWHAT⁴ and YOUR⁵, are purely chatbot-based, and they do not provide the ability to upload photos, and neither access to physicians.

Overall, MHEALTH applications have been developed to target a wide range of medical conditions. Most of the AI-based solutions deal with pimples, acne, scars, dark spots, pigmentation, and dark circles. On the other hand, the apps that deal with more serious health issues, such as skin cancer detection mainly rely on real physicians analysing the patients' inputs.

Dermatology is the branch of medicine that deals with diagnosing and treating skin disorders [13]. Developing MHEALTH apps targeting dermatological diseases is important for several reasons. First, such diseases can become highly contagious in environment that lack proper hygiene and medical attention, such as refugee camps. A notable example is the Syrian refugee camps in Lebanon, where medical support is mostly contingent upon non-governmental organizations (NGOs) involvement, especially with the ongoing economic crisis [14], the widespread of COVID-19 [15] and the 2020 Beirut's port explosion [16]. Second, some countries

³<http://www.sensely.com/>

⁴<https://medwhat.com/>

⁵<https://www.your.md/>

suffer from the scarcity of dermatologists per capita. For example, there exist roughly 11,000 skin care physicians in India,⁶ which translates to lesser than one dermatologist per 100,000 of the population, which makes it extremely difficult to cater for the needs of every citizen in a timely manner. Third, very few MHEALTH applications that specialize in dealing with dermatological infections.

CURESKIN⁷ is one of the few MHEALTH apps that specialize in dealing with dermatological diseases. It offers the option to upload images and automatically detects the rash or acne, without the need to consult with a physician. However, it lacks an interactive chatbot that collects personal information that proved to be useful to improve the quality of the diagnosis. SKINVISION⁸ is another application that deals with early cancer detection. This application uses the “fractal geometry” algorithm [17] to analyze uploaded images of skin grazes and moles. Furthermore, whenever available, a medical expert reviews the uploaded images to confirm or correct the preliminary diagnosis made by the app to.

Below, we describe recent research that was done to deploy machine learning algorithms for diagnosing different dermatological diseases. One approach, for melanoma recognition, combined deep learning with unsupervised sparse coding and Support Vector Machines (SVM) [18]. It used Caffe Convolutional Neural Networks (CNNs) [19] to extract features that are used to train the SVM classifier [20]. This system was trained on both RGB (red, green, and blue) and grayscale images, with the former providing better performances. The best performance was obtained when low-level features from both CNN and Sparse Coding were averaged prior to training the SVM. Another approach for melanoma recognition was explored in [21], where melanoma and non-melanoma images were segmented separately to boost the classification performance. Very deep neural networks (more than 50 convolutional layers) were used for both the segmentation

⁶<https://www.newzopedia.com/>

⁷<http://cureskin.com/>

⁸<https://www.skinvision.com/>

and the classification stages to obtain more discriminative features, and residual learning was used to overcome the degradation problem in very deep networks. Finally, the model developed in [22] used the VGG-M pre-trained neural network [23] to extract features that are then used to train a SVM classifier. The performance of the fine-tuned network was benchmarked on both the retina and melanoma datasets.

Dermoscopic pattern detection was carried out by training a CNN classifier that outperformed a strong baseline of SVM with a radial basis function (RBF) kernel [24]. As part of the preprocessing pipeline, input images were normalized by averaging the pixels' values and subtracting the mean. Also, data augmentation was performed by altering pixel intensities and geometric variations.

Unknown skin lesions were diagnosed in [25] using features that were extracted from the fine-tuned AlexNet: a CNN-based image classifier [26]. The fully-connected layers of AlexNet were converted into convolution layers, where the pre-trained weights from these layers act as convolution filters. Then, the extracted features were used to train a logistic regression classifier to classify skin lesions at 5 and 10 levels of class granularity. An updated version of the method altered the CNN feature extractor to become composed of multiple tracts [27]. Each tract considers the same image at a different orientation of itself, and the multi-resolution responses are then combined into a single layer using a supervised loss layer, thus making the final prediction a learned function of multiple resolutions of the same image.

The key goal of this thesis is to develop a MHEALTH system that relies on state-of-the-art machine learning algorithms to perform automatic diagnosis of selected common dermatological diseases that are widespread in refugees communities. The system architecture of this app is hybrid in the sense that it combines features extracted from a questionnaire that collects personal and medical information from the users, as well as features extracted from a state-of-the-art image processing model capturing information about the affected skin.

Chapter 3

System Description

In this chapter, we describe the system architecture of our MHEALTH system that is specialized in diagnosing dermatological diseases in vulnerable communities, mainly refugees, that lack clinical resources and healthcare support. This system can also help overcoming the imposed language barriers that prevents proper communication, leading to potentially inaccurate diagnoses.

3.1 System Architecture

The system architecture of the proposed MHEALTH system is depicted in Figure 3.1. This system was designed with the aim of mimicking an in-person appointment with a physician. At the high level, the system is comprised of a medical form (or survey) that targets its questions towards dermatology-related diseases and issues. The users' answers to these questions are converted into numerical features. It also allows users to upload images of the cutaneous lesions, which are embedded into a semantic space using a fine-tuned Convolutional Neural Networks (CNN). The resulting features are then fed to a machine learning model that outputs a prediction of the type disease. Finally, this information can then be sent to a specialist for approval.

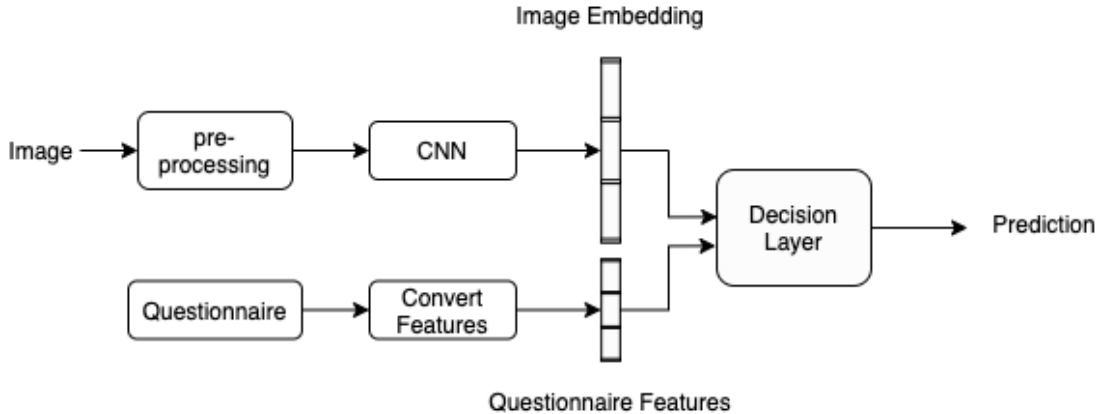


Figure 3.1: System architecture of our proposed MHEALTH method.

The proposed system was designed to have the following characteristics. First, it is based on filling out a medical form (or survey) in order to collect personal, medical and history information from the patient. Thus, it does not require direct and instantaneous contact with a physician. Second, the medical form can be easily translated to different languages, which is critical to overcome the language barrier between refugees and physicians that might prevent proper communications leading to potentially incorrect diagnoses or treatments. Finally, it is multi-source and multi-modal, in the sense that it is able to fuse information extracted from different data modalities and from different sources, particularly, (ii) the textual (from the medical form) and (i) the visual (the uploaded image) modalities. As a result, the machine learning classifier is enriched and trained with information collected from multiple sources, which should help boost its ability of perform accurate diagnosis.

3.2 System Modules

In this section, we describe in further details the different diagnosis modules, namely the *image-based*, *form-based*, *expert* and *mass* diagnoses, that are part of the proposed MHEALTH system.

3.2.1 Image-based Diagnosis

This module is formulated as an image classification problem, where it takes as input the uploaded images of the infected area of the skin, and outputs an embedded semantic representation that can be used to discriminate between different types of infections. This module is depicted in Figure 3.2. Once an image is uploaded to the system, it is pre-processed and then passed to the deep CNN model that is initially trained to predict the type of dermatological disease that is shown in the image. To train the CNN image classification model, we used DERMNET: a high-quality dataset that consists of 23,000 images for different skin diseases ¹.

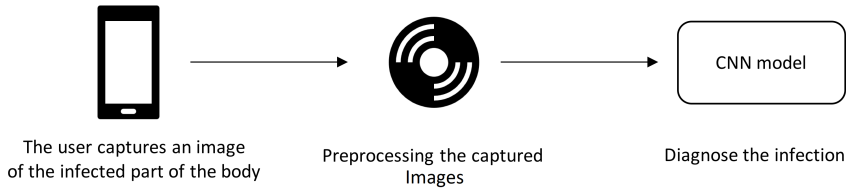


Figure 3.2: The image-based diagnosis module.

After acquiring the desired training images from DERMNET, preprocessing steps are applied to improve the performance and efficiency of the learning process. For instance, preprocessing ensures that the image orientation is random across all training and new data, to avoid bias towards a particular shape or orientation. Converting the images from a RGB (red, green and blue) scale to grayscale makes the learning process faster and less computationally expensive, as the model has to process only one-third the original data. Having said that, the performance is expected to do better with the RGB images as they contain more information. We evaluate this tradeoff later in Chapter 5. Filtering the images using smoothing filter helps in noise reduction by removing unimportant details such as hair and background noise. A sharpening filter can also help highlighting small details or blurred areas.

¹www.dermnet.com

Once preprocessing is done, the images are then used to train the image classifier, which is, in our case, an instance from the Convolutional Neural Networks (CNN) that are widely used for image processing.

3.2.2 Form-based Diagnosis

This module *asks* the users specific questions related to their personal history and their medical conditions or symptoms. In order to build this module, we have communicated with the Dermatology department at the American University of Beirut (AUB) to provide proper guidelines to come up with questions that are typically asked whenever patients visit their primary care physicians or dermatology specialists.

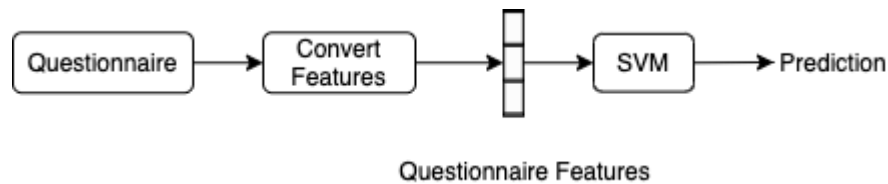


Figure 3.3: The form-based diagnosis module.

The module extracts the users' answers and transforms them into a set of features (as will be discussed later in Section 4.2.2) that can be either location-dependent, behavioral, or history-related. These features are then fed to a Support Vector Machine (SVM) model for prediction.

3.2.3 Expert Diagnosis

After obtaining the preliminary infection diagnosis provided by the above-mentioned modalities, this information can then be sent off to a professional medical doctor for approval or confirmation of the diagnosis. This step does not abrupt the automated diagnosing process, but it will further help optimizing and fine-tuning the trained models, by augmenting our training datasets with further features

obtained from the patients along with the approved diagnoses provided by the medical experts. This would improve the generalization capabilities of our system, in addition to its ability to learn newly-emerging diseases that were not part of the training dataset.

3.2.4 Mass Diagnosis

Finally, the obtained predictions are stored on a cloud server for further analysis. Considering the fact that refugees usually reside in settlements, the infections that one member has can spread to the whole community. The stored information can be analyzed to check for any trends about the spread of dermatological diseases, which can be extremely useful for tracing outbreaks and localizing their sources. It can also serve as an alarm system that detects the spread of infections before they turn into epidemics.

Chapter 4

Dataset and Models

4.1 Datasets

To train the diagnostic models that are described in Chapter 3, we used the following datasets: (i) a synthetic dataset to train the form-based model (dubbed as SYNFORM), and (ii) the DERMNET dataset to train the image classifier. Below, we provide a detailed description of these two datasets.

4.1.1 SYNFORM Data

This dataset is synthetically created by associating known features (symptoms and relevant information) with common dermatological diseases. These features reflect the kind of questions that are typically asked by dermatologists while examining each type of lesion. They were provided by a dermatology expert at the American University of Beirut Medical Center (AUBMC) after several rounds of discussions. These features are grouped under three different categories, as shown in Table 4.1: the (i) *behavioral* features describe how the lesion is affecting the patient’s skin, the (ii) *location* features describe which parts of the body is affected, and the (iii) *history* features describe family or personal history with certain symptoms or diseases.

Category	Features
Behaviour	<p>Itching, Itching worsens at night Lesions appear after itching Itching worsens during winter Itching better during summer Sweating makes itch worse White small scales on the scalp with variable itching Fibrosis of the papillary dermis Fever, Malaise, Headache Lesion becomes bigger over several days Muscle pain, Central chest Holes in nails Can get worse after a health event Can get worse after emotional stress Difficulty moving the joints Thickening in palms skin Thickening in soles skin Soles with orange-red coloration</p>
Location	<p>Face or elbows Genitalia Knees, torso, feet, neck or hands Lumbosacral area Nasolabial folds Popliteal Fossa Antecubital fossa The nipples, Scalp, Ears Body folds Medial portions of the eyebrows Upper eyelids Lateral aspects of the nose Retro-auricular areas The occiput Single skin lesion on the trunk Single skin lesion on proximal extremities</p>
History	<p>Family history of asthma Family history in atopic Family history of allergies Personal history of asthma Personal history of allergies Age</p>

Table 4.1: Summary of the different features that exist in the synthetic dataset.

Each instance in the dataset corresponds to an artificially created medical record that consists of a unique combination of features associated with a particular skin disease. We focus on instances that pertain to 6 types of skin diseases: *Atopic Dermatitis*, *Seborrheic Dermatitis*, *Psoriasis*, *Lichen Planus*, *Pityriasis Rosea*, and *Pityriasis Rubra Pilaris*. Each feature is represented with a value that ranges between 0 and 3, where 0 indicates its absence and 3 represents its maximum existence. It is worth mentioning that not all skin lesions have the same features, which eventually creates sparsity in the features matrix.

Each feature has a probability by which it is associated with a particular disease. For example, “oral lesions” are seen in up to 75% of patients with *Lichen Planus*. Therefore, we ensure that the “oral lesions” feature is activated for 75% of the *Lichen Planus* cases, independently from other features. This strategy allows to mimic how a real collected dataset could be, but also leads to limited amount examples with incoherent features combination. This can be considered as noise in the data, or outliers. The final SYNFORM dataset consists of 900 examples, and Figure 4.1 shows the distribution of the 6 classes in the dataset.

We can also notice that providing values for most of these features only require visual cues that can be perceived with the naked eye, in addition to a bit of knowledge about the patient’s and their family history. Therefore, users should be able to provide sufficiently accurate answers regarding these features, thus aiding the model’s ability to return the correct diagnosis. We can also notice the absence of important information that can be obtained via clinical testing and observations, which emphasizes the importance of the image-based diagnosis as a complementary source of information.

Next, we provide a detailed description of each of the 6 skin lesion diseases that will be considered in our experiments.

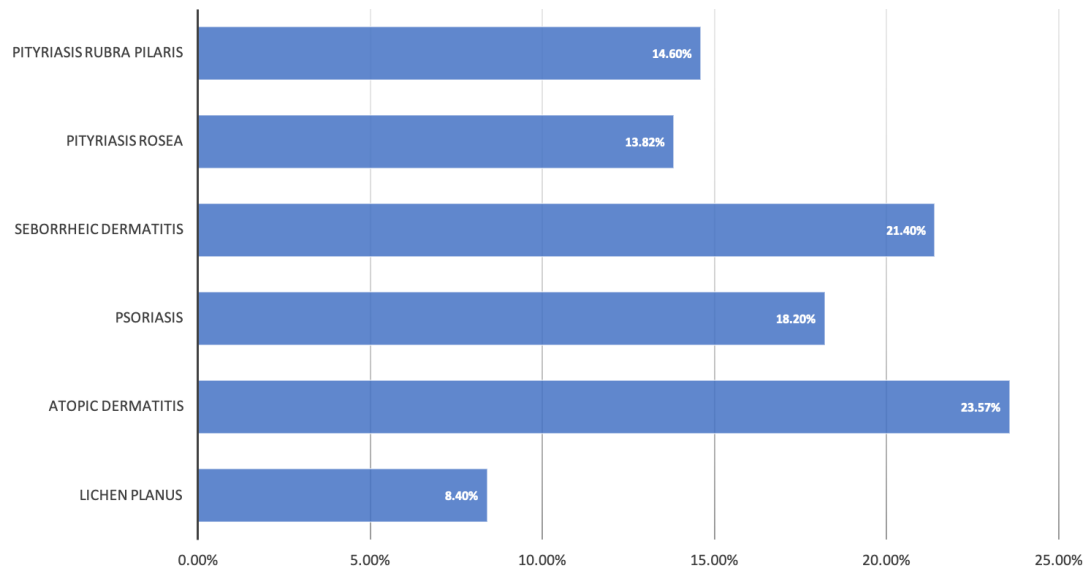


Figure 4.1: The distribution of the different classes in the SYNFORM dataset.

Atopic Dermatitis:

- * Itching is always present
- * Itching is worse in the evening/night
- * Skin lesions appear after the itch sometimes
- * Sweating and wool clothing makes the itch worse
- * Family history of atopic dermatitis is usually present
- * Family history of asthma and/or allergies increases the likelihood of having atopic dermatitis
- * Personal history of asthma and/or allergies increases the likelihood of having atopic dermatitis
- * In infants, the distribution of the skin lesions: face, elbows, knees, torso
- * In children/adults the distribution of the skin lesions: neck, hands, feet, popliteal fossa and antecubital fossa as well as the nipples.

Seborrheic Dermatitis:

- * Distribution of the scales and skin lesions mostly over the scalp, ears, face, central chest, and body folds
- * Worsens during the winter and better during the summer
- * White small scales on the scalp can be present with variable itching
- * On the face, skin lesions are symmetrical. They affect the forehead, medial portions of the eyebrows, upper eyelids, nasolabial folds, lateral aspects of the nose, retroauricular areas, and occasionally the occiput and neck
- * Chronic and usually comes in the winter and goes in the summer.

Pityriasis Rosea:

- * It usually starts with a single skin lesion on the trunk or less commonly over the neck or proximal extremities and the lesion becomes bigger over several days.
- * 5% of the patients have fever, malaise, muscle pain, and headache.
- * Several lesions then begin appearing all over the body especially the trunk and the back
- * The skin lesions have scales especially paralleling the perimeter
- * The face, palms, and soles are usually spared.

Psoriasis :

- * The risk in the general population is 1-3%; if a single sibling has it, then the risk increases to 6%, if one parent has it, the risk becomes 16% and if both parents have it then the risk is 41%.

- * Psoriasis has changes in nail appearance (holes in nails) in 79% of patients.
- * Psoriasis skin lesions can appear in locations that have been subjected to trauma (fall or injury)
- * Lesions can get worse after a health event or emotional stress (sickness, hospitalization, loss of a family member, etc)
- * The scalp, elbows, knees and lumbosacral area (lower back/buttock) are sites of predilection, as are the hands and feet.
- * Genitalia is a common affected area in up to 45% of cases.
- * Itching and pain can be present but not always.
- * Depending on the studies, 5-30% of patients with psoriasis might have hand joints problems including pain and difficulty moving the joints especially after waking up.

Pityriasis Rubra Pilaris:

- * Shows no gender bias, affecting both men and women equally,
- * The incidence has two peaks. The first is during the first and second decades and the second is during the sixth decade.
- * Lesions may start in the head and neck and progress caudally; alternatively, they may involve the entire body.
- * The nails might be involved: thickening of the nails and debris underneath the nails is a hallmark.
- * A form of Pityriasis Rubra Pilaris can be associated with the HIV infection
- * Involvement of the palms and soles is sometimes seen thickening of the skin of the palms and soles with an orange-red coloration.

- * If very early, Pityriasis Rubra Pilaris can mimic seborrheic dermatitis
- * Scales, if present, are finer than in psoriasis.

Lichen Planus:

- * Some studies found that women were affected nearly twice as often as men.
- * Has been reported to affect from 0.2% to 1% of the adult population
- * Has its onset during the fifth or sixth decade, with two-thirds of patients developing the disease between the ages of 30 and 60 years.
- * Oral lesions are seen in up to 75% of patients with skin Lichen Planus.
- * Occurs in up to 10% of first-degree relatives of affected patients.
- * In several case-control studies, the prevalence of HCV (hepatitis C-virus) (3.5–38%) was 2- to 13.5-fold higher in patients with Lichen Planus than in controls.
- * Of the various types of Lichen Planus, it is the oral form that is most commonly viewed as a manifestation of HCV infection. By PCR, HCV RNA (RNA is the genetic material of HCV or hepatitis C virus) was detected in 93% of oral Lichen Planus lesions
- * Lesions can also occur at the site of trauma
- * The most common sites of involvement are the flexor wrists and forearms, the dorsal hands, the shins, and the presacral area. Mucous membranes, especially the oral mucosa (see below), are affected in more than half of patients, and this is often the only site of disease.
- * Lesions are always itchy with minimal to no scaling.

4.1.2 DERMNET Data

DERMNET is one of the largest publicly available databases in dermatology imagery, with more than 23,000 images covering a wide variety of skin diseases. This dataset is organized biologically in a two-level taxonomy. The low-level has more than 600 skin diseases, whereas the high-level contains only 23 skin disease classes, where each of these high-level classes contains a sub-collection of the low-level classes.

We are not planning to use all these images, since they are not all relevant to our problem, which is developing a MHEALTH system that helps diagnosing skin diseases among refugees. In fact, not all of the available skin lesions are commonly spread among refugees. Recent studies¹ have shown that the most spread skin diseases are *Giardiasis*, *Leishmaniasis*, and *Echinococcosis*. Other studies reached the conclusion that common skin diseases in refugees camps are infections caused by poor hygiene and overcrowded living spaces, such as *Scabies* and *Pediculosis*².

Nevertheless, given the size of *DermNet* dataset, we decided to pick the skin lesion classes that were well-represented in the target audience of the system. Also, since the form-based classifier is used to complement the image-based classifier, they must both be trained with datasets that are consistent in terms of the output classes they contain. Therefore, we will only consider the six classes that already exist in the SYNFORM dataset. This reduces the dataset to 4,098 images. Figure 4.2 illustrates the distribution of the 6 classes in the dataset.

¹<https://www.cdc.gov/immigrantrefugeehealth/profiles/syrian/health-information/parasitic-infections/index.html>

²<http://www.derm.city/single-post/2016/06/01>

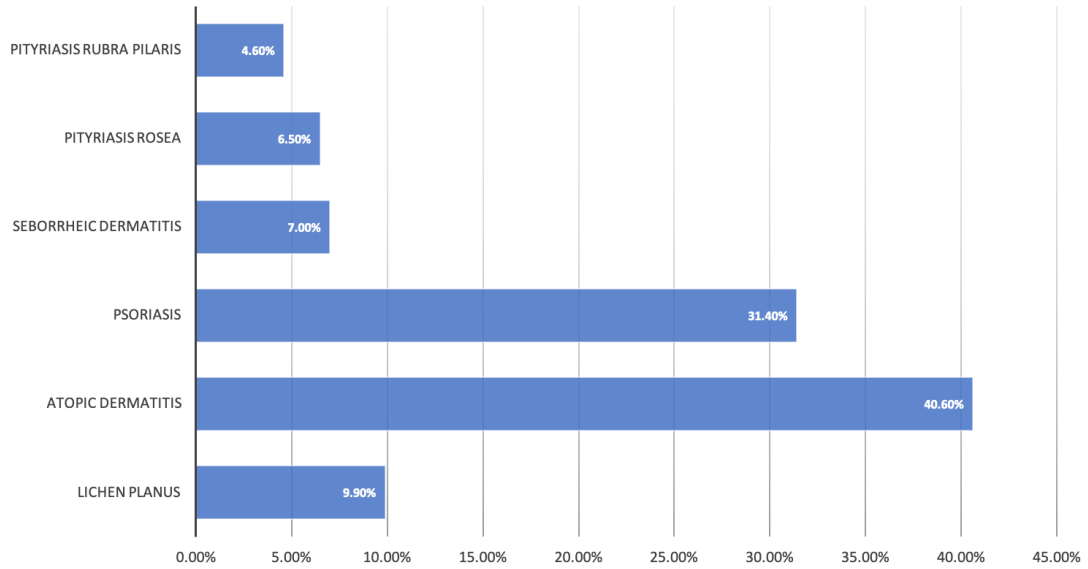


Figure 4.2: The distribution of the different classes in the DERMNET dataset.

4.2 Data Transformation

In order to obtain the optimum performance out of our models, certain pre-processing or transformation techniques need be applied to each of our training datasets.

4.2.1 Images: DERMNET

Several preprocessing techniques can be applied to images to help boost the performance of the fine-tuned model. At first we applied a smoothing filter to remove small details that are irrelevant to the task at hand, then we implemented a sharpening filter to increase the details on the remaining information in the image that might be useful.

It is important to discuss the effects of transforming the images from colored (RGB) to grayscale. This transformation represents a well-known tradeoff between accuracy and efficiency. While it might lead to a decrease in accuracy because of the loss of potential information that exist in the dropped colors, this

transformation improves the model’s efficiency as it significantly reduces the input data that needs to be processed. Intuitively, choosing to keep the images in their colored form seems obvious as they contain more information represented in the different color gradients of the skin and the lesions. But, considering that we do not exactly know if these extra information are going to be useful or not in our case, we decided to fine-tune the models using both colored and gray-scaled images, and compare the difference in the resulting performances.

The images’ distribution over the selected classes, as shown in Figure 4.2, reveals a clear data imbalance. Therefore, data augmentation techniques become useful to balance the dataset, and consequently to improve the model’s performance. We applied image rotation over 90, 180, and 270 degrees. We also applied change in brightness, mirroring, random occlusions, and shearing.

The final step involves cleaning the dataset from anomalies, which is critical for the learning process. The DERMNET dataset has a lot of examples that either are either microscopic images, which are useless for our system, or images that capture the whole body while the lesion is on a small part of it. Such images were excluded from our dataset.

4.2.2 Form: SYNFORM

When it comes to categorical data representation, we often suffer from the issue of sparsity, which typically hurts the learning process and decreases the effectiveness of the resulting model. We need to handle this issue, given that our data that is coming from the answers to the system’s form has the potential to be highly sparse, especially if represented using one-hot encoding. One common way to avoid this problem is through feature reduction algorithms, such as Principle Component Analysis [28]. Despite the fact that PCA helps getting rid with the sparsity problem, we did not apply it in our system because it might possibly lead to unnecessary loss of information during the data reduction process.

Another way to handle sparsity is to perform Entity Embedding [2]. This technique maps a given feature set to a new one that has a smaller number of dimensions. For instance, a feature that can take one of four distinct values can be represented using two features, instead of four features when one-hot encoding is used. Furthermore, entity embedding converts categorical values to soft and dense values which are more favourable to classifiers, and can also reveal the potential properties of a variable. It also improves the model’s generalization capabilities by having a much higher set of categories for a single feature. Additionally, it reduces memory usage and speeds up training compared to one-hot encoding representation.

This is done by training a neural network with an embedding layer by assembling the categorical values that belong to the designated lesion. This representation allows us to obtain inherent properties from each categorical value. The output is then used to replace one-hot encoded values. Further details about the Entity Embedding model is available in Appendix B.

4.3 Image Classification Models

Recent advances in Computer Vision lead to the development of a huge number of Convolutional Neural Networks (CNNs). Many CNN networks architectures were developed in order to reach the highest performance possible on the ImageNet [29] dataset which is considered the benchmark for the image classification task.

CNNs can be used to perform image classification in our MHEALTH system under different settings. One option is to train a CNN model from scratch with only dermatology-specific images, which is only a good idea if we have an abundance of skin lesion data. Another option that is most commonly used to get a more accurate learning is to initialize our model with one of the most recent state-of-the-art pre-trained models, and then fine-tune it with our dermatology-specific dataset. It is worth mentioning that using pre-trained models, despite

them being trained on different data, is highly recommended as the model already knows how to detect various borders and features even for unseen data. An additional preprocessing step would be to either crop or resize the image to the size of the network’s input. The authors in [30] fine-tuned the the VGG16 and VGG19 networks [23], proving the benefits of training on pre-trained models.

While there exists a large number of pre-trained CNN models, our model selection was based on several criterion, including the model size (number of parameters) and the reported performance. The number of parameters in the selected models is always a limitation due to the fact that more parameters mean more memory and time requirements for training and storage. For instance, the top performing model, `FIXEFFICIENTNET-L2`, has over 480 million parameters. Finding a server big enough to fine-tune this model is both expensive and time consuming. It is also worthy to compare different CNN architectures that had high and close reported scores on ImageNet, and how they performance on our data. Examples of different CNN architectures that will be used in our experiments include: the `INCEPTION` models, the `RESIDUAL` Network models, and the `EFFICIENTNET` models. Next, we list instances of the selected CNN models with a brief description on their architecture and score on ImageNet.

Inception3 This model was proposed by [31] and has 24M parameters. Its architecture, and generally all inception architectures are aimed to factorize convolutions to reduce the number of connections and parameters without losing efficiency. This has been done by factorization convolutions into smaller ones, for instance, a 5x5 (total of 25) convolution is replaced with two 3x3 (total of 18) convolutions. They also used asymmetric convolutions, as in replacing 3x3 (total of 9) convolutions with a 1x3 and a 3x1 (total of 6) convolutions. As we observe, by doing these factorization steps, the complexity of the model has been maintained while reducing the number of parameters and by definition, reducing the risk of overfitting.

Xception This model was proposed by [32] and has 22.8M parameters. It stands for Extreme version of Inception. This model introduces a modified depth-wise separable convolution which is basically pointwise convolution followed by a depth-wise convolution. A pointwise convolution is a 1×1 convolution to change the dimension applied across the channels (the depth across the image channels). Whereas depth-wise convolution is a channel-wise $n \times n$ spatial convolution. For example, assuming we have 5 channels input, then we will have 5 $n \times n$ spatial convolution. Comparing this with the conventional convolution, there will be no need to perform a convolution operation across all channels, which means less connections, hence a lighter model.

Inception4 This model was proposed by [33] and has 43M parameters). This inception model is intended to make a more uniform inception3 factorization modules. This can enable us to boost performance by adding more of these uniform modules.

InceptionResNetV2 This model was proposed by [33] and has 55.8M parameters. It is a hybrid model between Inception modules and Residual Networks (ResNets), that help solve the degradation issue (more layers leading to slower convergence or none at all) by adding the input of hidden layer (or several layers) to their output, thus learning the residuals. The InceptionResNetv2 model corresponds to the Inceptionv4 modules with residual connection across each module.

EfficientNet-B7 Noisy Student This model was proposed by [34] and has 66M parameters. It is based on understanding the effect of scaling the network in different dimensions. The conclusion reached is that balancing all networks dimensions depth, width, and image resolution, performed the best increase in overall performance. This EfficientNet has been trained using the Noisy student algorithm. First you train a "teacher model", which is later used to label unlabeled

beled images. To ensure accurate enough predictions, a threshold has been set to filter out less confident predictions. This data is now merged with the labeled data and "student model" is trained on this combined data. Now, the student model becomes the "Noisy Student" as it is trained on data labeled from the teacher model, and is used to label more of unlabeled data. The loops goes on until convergence.

Table 4.2 displays a summary of the discussed CNN models presenting number of parameters and score on the ImageNet dataset for each architecture.

In the next chapter, we discuss how the above-mentioned models are used in different configurations and preprocessing schemes in order to obtain the best possible performance.

Model	Number of Parameters	Score on ImageNet
Inception3	24M	78.8%
Xception	22.8M	79%
Inception4	43M	81.2%
InceptionResNetV2	55.8M	83.1%
EfficientNet-B7 (Noisy-Student)	66M	86.9%

Table 4.2: CNN models specifications.

Chapter 5

Experiments and Results

In this chapter we describe details of the experimental setup used in this thesis. This includes the different configurations under which the selected CNN models were trained, namely *standalone* and *majority voting*, with or without the features extracted from the form. We also describe the metrics that we used to evaluate the performance of the different models. Then, we present the results of each configuration and perform a qualitative analysis to draw conclusions. The SYNFORM synthetic dataset was used in our experiments to train and evaluate a classification model that would serve as baseline to compare the novelty detection process with.

As we move forward, we notice that the count of the DERMNET dataset surpasses the count of SYNFORM data significantly. This discrepancy in the data, and considering that both datasets come from different sources, compelled us to randomly pair instances from the two datasets that belong to the same class. Although some instances from the SYNFORM may match their DERMNET counterpart, but the likelihood that the pair will not match across all the features is way higher especially visual features such as age, and location. This issue is expected to hurt the confidence in the models' performances and can only be fixed by properly collecting data from hospitals and sources that have and are

willing to share data with similar characteristics to the ones we are using in this thesis.

5.1 Experimental Setup

5.1.1 Baseline

First, we identify a strong baseline to compare and justify the final system architecture and the proposed features, and in order to quantitatively assess the suggested solution. Given that we are using two datasets to train our models: SYNFORM to train the form-based model and DERMNET to train the image classifier, our experimental setup is driven by the following goals. First, we need to showcase that the features extracted from the form are well-defined, standalone, and can be used to learn to predict their corresponding labels, which in our case are the 6 pre-defined skin diseases. In other words, we need to see that training a model using only these features should yield sufficiently good performances that reflect their predictive capabilities. The resulting performance of this configuration is considered the baseline, to which further configurations will be compared with to evaluate their effectiveness. Second, we need to prove that the importance and the added value of the image-based model in terms of its ability to provide complementary information to the form-based model (baseline), thus boosting the results to new levels.

The form-based model corresponds to a Support Vector Machines (SVM) classifier that is trained with the features existing in the SYNFORM dataset. At inference time, those features will be collected from the users once they answer the questions in the form. To get the best performance from the SVM model, we used the grid search algorithm to fine-tune the model’s hyper-parameters: the kernel’s type, the kernel’s width σ (for RBF), and the misclassification parameter C . The *gamma* parameter is a parameter for the Radial Basis Function (RBF) kernel

that defines the influence of each instance in the training data. Small values of γ indicate a large similarity radius, resulting in more instances being grouped jointly, whereas larger values means that the instances need to be closer to one another to be considered in the same class. The parameter C penalizes each misclassified data instance, and thus helps find the balance between increasing the distance of the decision boundary to the classes (the support vectors) and maximizing the number of instance that are correctly classified in the dataset.

Since our dataset contains 6 classes (or labels), we trained a multi-class SVM classifier using two different approaches. The first approach is the “One-vs-All” (OvA), which is a heuristic procedure that uses binary classification for multi-class classification. This is done by splitting the multi-class dataset into multiple binary classification problems, and a binary classifier is then trained on each binary classification split. For example, to train a model that detects the *Psoriasis* cases, we re-label the dataset examples as *Psoriasis* or *non-Psoriasis*, and the same applies for each of the remaining classes. The other approach is the “One-vs-One” or (OvO), which is another heuristic approach that splits the data into one binary dataset for each class in the dataset. In this case, in comparison to the OvA example, the *Psoriasis* instances are going to be trained against each of the other classes separately, and so on.

5.1.2 CNN Standalone

To perform image classification, we trained the EfficientNet-B7 (Noisy-student) as a standalone model without any added features. This model is trained using both RGB and grayscale images, to ensure fair comparison to the other approaches. This model was selected to be trained alone based on its best performing score in a set of preliminary experiments.

5.1.3 Majority Voting

The main incentive for using the “majority voting” strategy is to discover a hypothesis that is not necessarily contained inside the hypothesis space of each of the various models from which the ensemble is built. Experimentally, ensemble methods tend to yield better results when there is a significant diversity among the trained models. This approach is represented in Figure 5.1. Hence, we decided to apply the majority voting on a collection of pre-trained and fine-tuned CNN models that have different architecture. We applied the ensemble mechanism under two scenarios. (i) *Only CNN models*: in this case we trained the Xception, Inception4, and the InceptionResNetV2 models using both RGB and grayscale images, and (ii) *CNN with form features*: in this case, the SVM model that was trained using the features from the SYNFORM data (the baseline model) is added to the mix of CNN models, then majority voting is applied to the outcome.

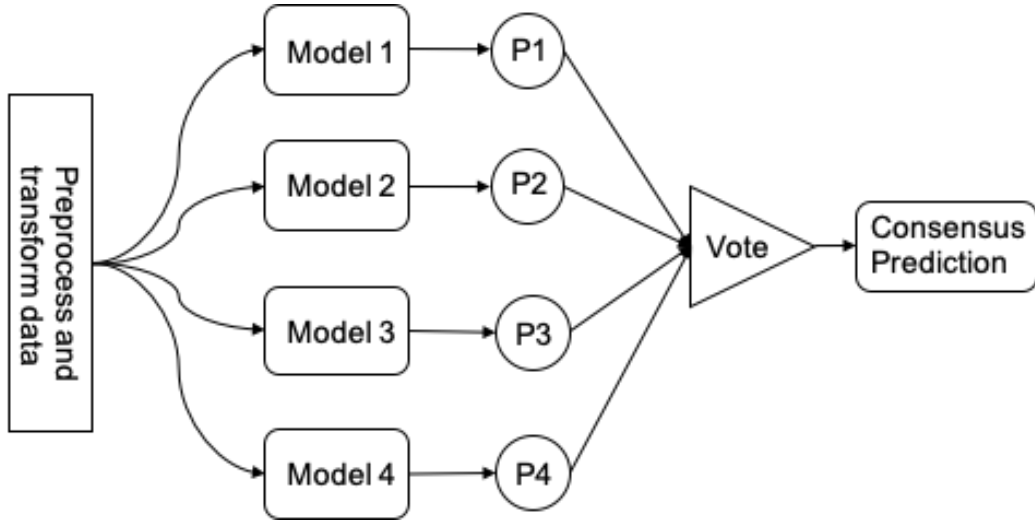


Figure 5.1: Majority Voting approach structure.

5.1.4 Fusion: CNN with Form Features

Instead of adding the SVM next to the CNN in an ensemble setting, we concatenated the Form Features to the image features extracted from the first fully-connected layer following the series of convolutions in the CNN. In other words, we concatenate the Form features to the image representation that is produced by the CNN model. This experiment is also conducted on both RGB and grayscale images, in order to ensure fair comparisons to the other configurations. This architecture is displayed in Figure 5.2

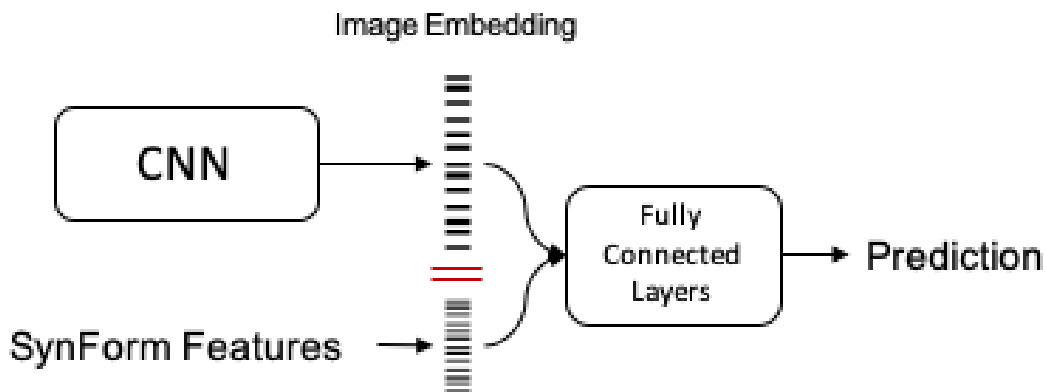


Figure 5.2: Feature Fusion approach structure.

5.2 Evaluation Metrics

In order to substantiate the fidelity of our results, we use a set of metrics, based on which we can fairly and accurately compare and validate the performance of each of the previously discussed scenarios.

Student's t-test Dealing with neural networks means dealing with random weights initialization. This means that every time we train a CNN model we might get different results. These results might be close to or distant from each other, and this behaviour depends on the distributions of the training and testing

data, as well as the weights initialization procedure. If the difference in the results of the same model across multiple runs is high, then it is not accurate to claim that any of these results are valid without performing a *t-test* first. The *t-test* provides an indication of the significance of the difference between two (or more) sets of results. This allows us to compare the performance of our models more accurately without risking being affected by an outlier model. Each model is trained using 10-fold crossvalidation with random data to insure that the data across the different iterations are independent. The *t-test* is calculated as shown in Equation (5.1).

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(S_1)^2}{n_1} + \frac{(S_2)^2}{n_2}}} \quad (5.1)$$

where \bar{X}_i is the mean of the i^{th} model, S_i is the standard deviation of the i^{th} model, and n_i is the number of runs of the i^{th} model.

Recall This metric is a good measure to determine, when the costs of False Positives is high. It answers a very simple question: how many of the predicted instances were predicted positive out of everything that is originally positive? This measure is of interest to our MHEALTH system as it puts more emphasis on the ratio of correctly-identified lesions to patients who belong to the same lesion classes. Equation (5.2) illustrates how recall is calculated,

$$recall = \frac{t_p}{t_p + f_n} \quad (5.2)$$

where t_p is the number of true positive predictions, and f_n is the number of false negatives predictions.

F_1 Score This metric is commonly used in the case where the training and evaluation datasets exhibit class imbalance. This metric provides a balance between

Recall and Precision (which tells you what portion of the actual positives was predicted correctly). Equation (5.3) illustrates how the F1 score is calculated,

$$F_1\text{score} = \frac{2 * t_p}{2 * t_p + f_n + f_p} \quad (5.3)$$

where f_p is the number of false positive predictions.

5.3 Experimental Results

In this section, we demonstrate the results of our proposed methods. Section 7.1 displays the results of the hyperparameters fine-tuning for the four supervised classification models. Section 7.2 discusses the achieved results of the classification and the value of the novelty prediction process.

We start with setting up the baseline model, which will be used to identify the advantages or disadvantages that other configurations and approaches bring to the table. We used the SYNFORM data to train a SVM classifier. We used two types of kernels for the SVM model: the Polynomial kernel and the Radial Basis Function (RBF) kernel. The grid search algorithm was used to fine-tune the hyperparameters C and the RBF kernel’s width $gamma$, where preliminary results indicated that tuning the C parameter has no significant impact when using the polynomial kernel.

Results are illustrated in Table 5.1. It can be observed that a C of value 0.001 and a $gamma$ value of 0.2 on a RBF kernel achieved the highest results of 82.41 F_1 score and 83.12 recall.

Regarding the Majority voting approaches, we begin with training the different CNN models, individually. Considering memory limitations, we used a batch size of 16. We also set the dropout rate of 0.75. This is a common approach to improve the model’s ability to generalize to new unseen images. Table 5.3 illustrates the results of the different CNN classifiers on both color modes.

Kernel	Hyper-parameters	F1 score	Recall
Polynomial	degree = 3	79.1	79.23
	degree = 4	80.36	79.88
	degree = 5	78.18	76.93
RBF	$C = 0.001 \quad \gamma = 0.2$	82.41	83.12
	$C = 0.1 \quad \gamma = 0.2$	79.36	77.57
	$C = 1 \quad \gamma = 0.2$	76.18	75.62

Table 5.1: Results of the baseline model that only uses the “Form” features.

CNN model	Color mode	F1 score	Recall
Xception	Grayscale	72.84	72.03
	RGB	77.57	77.24
Inception4	Grayscale	71.28	70.72
	RGB	78.97	78.86
InceptionResNetV2	Grayscale	74.48	74.91
	RGB	79.29	80.57

Table 5.2: The individual performance for the different CNN classifiers.

Results confirm that converting to grayscale significantly hurts the performance, and that the InceptionResNetV2 achieves the highest results.

After acquiring the individual results for each model, we can proceed with the majority voting process. First, we evaluate the ensemble using only CNNs as displayed in table 5.3. Then, we add the Form SVM model (the baseline) to the ensemble mix, and compare performance. Results in table 5.5 indicate that incorporating the Form-based model consistently improves the results, which confirms that the Form features are complementary to the images.

Finally, we evaluated the impact of adding the Form features to a strong standalone CNN, namely EfficientB-7 (Noisy student). To incorporate the form features, we concatenated them to the image embedding generated by the CNN

Form features	Color mode	F1 score	Recall
No	Grayscale	74.89	72.29
	RGB	80.71	79.64
Yes	Grayscale	82.14	82.76
	RGB	83.23	83.78

Table 5.3: The performance of the ensemble CNN model with and without incorporating the Form-based model.

model. Then, fully connected layers and a softmax classification layer were added on top of the fused vector to perform classification. Results in Table 5.4 confirm that incorporating the Form features to the image features boost the model’s ability to perform accurate diagnosis. However, the increase in performance (1.7 points growth in F_1score) was not as significant as one would expect it to be, given the richer context being provided. This can be mainly attributed to the way the SYNFORM dataset was created, which involved randomly assigning instances from the dataset to DERMNET instances of the same class (refer to Chapter 5). While this procedure would obviously lead to confusing the system during training, yet the increase in performance indicates that it was able to make the model focus on more important features, while ignoring those hurting the learning process.

In the next section, we will discuss in details what meaning these result hold and how we can benefit from this feedback to further improve the performance.

5.4 Results Analysis

After obtaining the experimental results, it is important to understand what they actually mean and how we can further reduce the error rate. One expected

Form features	Color mode	F1 score	Recall
No	Grayscale	79.52	79.03
	RGB	84.67	86.24
Yes	Grayscale	80.18	77.36
	RGB	86.35	86.95

Table 5.4: The performance of the EfficientNet-B7 (Noisy Student) model with and without the Form features.

observation was that all experiments that were conducted using the grayscale images yielded inferior results compared to their RGB counterpart. This is mainly due to the loss of information, as color depth and shade contain more information especially in skin lesions detection where different shades of the lesions can hold different explanations. Therefore, we exclude experiments on grayscale images from further analysis.

As for the majority voting approach, we noticed a 2.5% increase in performance (improving from 80.7% to 83.2%) on F1 score. To better understand how this happened, we need to check the performance of the models per class separately.

First we start inspecting the models performance per class as we show in table 5.5. We can notice from the results that the SVM performance is more balanced than the ensemble. A great factor of why the ensemble is lacking behind on some classes, is the unbalanced distribution of the symptoms of the instances per class. Meaning, some of the images that we collected cover a lesion with mild symptoms that are not clearly visible compared to the other images in the class as is seen in Figure 5.3.

Another major factor is the color of the skin. Images where lesions appear on a dark skin have different characteristics than when they appear on a lighter skin. Difference in color between damaged areas and unaffected areas are more



Figure 5.3: Comparison between cases with a) mild and b) visible symptoms [1].

Skin Lesion	SVM	CNN ensemble
Atopic Dermatitis	85.67	85.35
Lichen Planus	86.74	84.20
Pityriasis Rosea	80.37	76.83
Pityriasis Rubra Pilaris	80.92	74.13
Psoriasis	83.80	86.45
Seborrheic Dermatitis	80.56	79.53

Table 5.5: F_1 score per class comparison between SVM and CNN ensemble

visible on white skin than black skin as depicted in Figure 5.4. Added to this, the examples with black skin are fewer in numbers than examples with white skin. This is why data augmentation did not add much value to solving this problem in particular.

Moreover, the images are taken in different angles. In some cases the lesion covers a large area of the body, this also leads to taking a picture that covers all of damaged areas as we can see in Fig 5.5. Inputs like that are not as common as



Figure 5.4: Comparison between a symptoms on white skin (a), and symptoms on dark skin (b) [1].

the closeups images in the dataset which makes it harder for the model to learn and predict these cases.

These discrepancies add to the complexity of the data. These complexities mostly come from unbalanced distribution of cases within each class whether in color skin, severity of the symptoms, the size of the covered area of the body.

As for the last approach, we display in Table 5.6 results of training EfficientNet-B7 (Noisy Student) as a standalone model in comparison to adding the form features to the decision making layers.

We noticed how classes that were behaving badly on CNNs such as Pityriasis Rosea and Pityriasis Rubra Pilaris have improved on the EfficientNet-B7 model. They have also improved more than other classes in the data after including the Form features to the model. These results show the added value of the added features in increasing the performance over the classes that the CNN alone had trouble predicting. In other words, the decision layers were able to realise that giving more weight to the added features improves the capability of correctly



Figure 5.5: Example of an image with large covered area of the body [1].

Skin Lesion	EfficientNet-B7	EfficientNet-B7 + Form
Atopic Dermatitis	86.14	87.36
Lichen Planus	84.20	85.06
Pityriasis Rosea	82.47	84.88
Pityriasis Rubra Pilaris	81.32	84.30
Psoriasis	86.21	87.35
Seborrheic Dermatitis	83.61	85.71

Table 5.6: F_1 score per class EfficientNet and EfficientNet with form features

predict weak classes.

Chapter 6

Conclusion and Future Work

In this thesis, we have designed and implemented a novel classification system to perform early diagnosis of dermatological diseases. Such a system can be useful in societies that lack robust access to medical care services, such as refugees camps.

The inputs to this system are presented in the form of images capturing the harmed segment, in addition to information obtained from the patients answering a specifically-designed questionnaire. The inputs are processed according to their type. In particular, the main contribution of this thesis lies in the novelty of fusing the multi-modal features together to perform classification. We trained and evaluated several supervised machine learning classifiers, including Support Vector Machines (SVMs) and Convolutional Neural Networks (CNNs) with different input features, and we compared them in terms of accuracy. To this end, we evaluated post-inference model-level fusion through ensemble mechanism, vs. feature-level fusion by concatenating extracted features prior to performing classification. The best performing system was achieved through feature-level fusion, by concatenating features extracted from the questionnaire with the embedded representation of the input image obtained via the EfficientNet-B7 CNN model, and then performing classification with a logistic regression layer.

In order to build this framework, we collected training data from two sources that were not considered or were not publicly available. The images were crawled and gathered from a website used by dermatology students. Whereas the questionnaire data was synthetically created with the help of a dermatology resident at the American University Hospital. Finally, the generated dataset depict the main part of the first deliverable.

In this work, we faced limitations presented by the lack of datasets in the research community concerning non-cancerous dermatology data. Additionally, the distribution of the collected data was out of balance concerning skin color and the lesion progression level. We consider the latter limitation out of this work's scope because it is extremely difficult to collect enough data manually and the procedure of excluding the unwanted images will take a lot of time considering the data is already filtered. Additionally, it will decrease the data size of some classes that already are under-represented compared to other classes.

A considerable setback in this work was the lack of data for the SYNFORM. The synthetically generated the data lead to mismatching the SYNFORM instances with the images dataset. Although this situation created more room for the model to generate error during prediction, it allowed us to see that in the worst case scenario the performance increases compared with the standalone models. A major direction for future work is to collect data from medical sources where they are able to provide images alongside their SYNFORM counterpart. From our findings during this thesis indicates this coherent dataset will definitely add further significant improvement.

This collection should also be balanced throughout skin colors. As we realized through our experiments, diagnosing diseases over black skin was lacking in accuracy and a big reason why that happened was due to little provided data for non-white skinned images. It is also worth mentioning that even physicians find it more difficult to diagnose people with darker skin, so it is definitely an important part of the research to be included.

Lastly, the system that was built in this thesis needs to be implemented in an interactive platform, where users can insert images and fill in forms in order to get diagnosed. The collected data should also be sent, after consent, to a physician to confirm the system's diagnosis or correct it. This new data is critical to further train and improve the overall system.

Appendix A

Smartphone Sensors for mHealth

In this appendix, we list the different sensory systems that have been implemented in smartphones for mHealth applications.

Sensors	Domain	Applications
Camera	Photo and Video capture	It was used to track different diseases, to view surgical effects, for remote diagnostics, incision monitoring, skin disease analysis [35], and to supervise children's health [36].
GPS	Location tracking	It provides access to track vulnerable patients such as elderly, people with Alzheimer disease [37] and victims of Ebola [38] by using contact-tracing applications.

Electro-cardiograph	Cardiovascular disease monitoring	Electrocardiograph-enabled mobile phones were used in underdeveloped areas in china for surveillance of heart diseases [39].
Wi-Fi	Data sharing and communication	The Wi-Fi module empowers the smartphone to communicate the health data to physicians for diagnostic and treatments.
Bluetooth	Data sharing and Communication	Enables the short-range data communication between mobile phone and various health monitoring devices and wearable sensors.
Microphone	Voice recording	Provides the communication with physicians regarding diagnostic and clinical support. It also provides capital for the audio analysis to access the patients' feelings with different diseases such as myotonic syndrome [40].
Accelerometer	Acceleration measurement	Helps to measure the device's orientation relative to earth and to estimate the motion. It can be used to monitor gait and step counting which can help in early diagnosis of Parkinson [41].

GPS, accelerometer, compass, gyroscope, barometer	Physical activities	The combined module is exploited for measuring the sedentary versus non-sedentary activities.
Microphone, accelerometer, GPS	Social engagement	This package enables the surveillance of mental health by monitoring the social encounters, conversationalist talks, anxiety, stress depressive behaviors and crustal motion of patients [42].
Microphone, GPS, touch, accelerometer, interface, light sensor	Sleep Pattern tracking	This module provides the effective information of disrupted versus continuous sleep patterns of a patient [43].

Table A.1: Overview of the capabilities of the different smartphone sensors for mHealth applications [3]

Appendix B

Machine Learning Models

In this appendix, we provide a brief description of the main machine learning models that will be used in this thesis.

B.1 Support Vector Machines (SVM)

A discriminative machine learning classifier that operates by separating instances from different classes using a hyperplane [20]. In other words, given a training dataset, the model produces an optimal hyperplane that that can best separate the given classes. The hyperplane is defined by its margins, where the best classification is reached when the distance to these margins from the nearest data points (also referred to as support vectors) is maximized, as shown in Figure B.1 This is equivalent to minimizing the weights as shown in Equation (B.1).

$$\min_{w \in R^d} ||w||^2 + C \sum_i^N \max(0, 1 - y_i \cdot f(x_i)) \quad (\text{B.1})$$

Classifying data with multiple dimensions using a hyper plane is also made possible by using kernels that transforms the data to a higher dimension representation to the point that they can be separated, and a hyperplane is able to separate the classes.

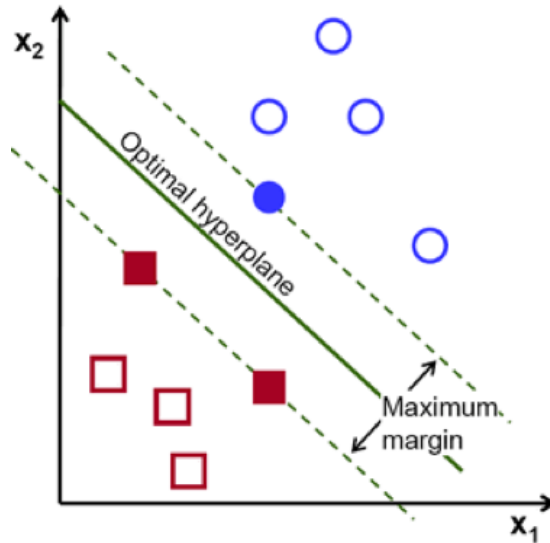


Figure B.1: The maximal marginal hyperplane in SVM. source:

<https://www.fabienplisson.com/svm-model/>

B.2 Convolutional Neural Networks (CNN)

One type of Artificial Neural Networks (ANN) that try to mimic the human vision; scanning an image and extracting features from different levels of abstraction to identify the objects in the image [9]. Although CNNs are originally developed and used in computer vision, they have also been successful in other domains, including NLP [44].

The architecture of a CNN model is composed of a sequence of layers from different types; convolutional, rectified linear unit (ReLU), pooling and fully-connected (FC) layers, as illustrated in figure B.2.

The convolution layer extracts features from the input image as it preserves the spatial relations between pixels by learning image features using custom filters that are convolved over the image. An example of a filter is shown in figure B.3. One or more filters K_i are convolved with the whole input image I to create one or more feature maps $I * K_i$ that identify simple objects (curves or edges) at the lower levels of the CNN, and more abstract objects (faces, animals) at higher

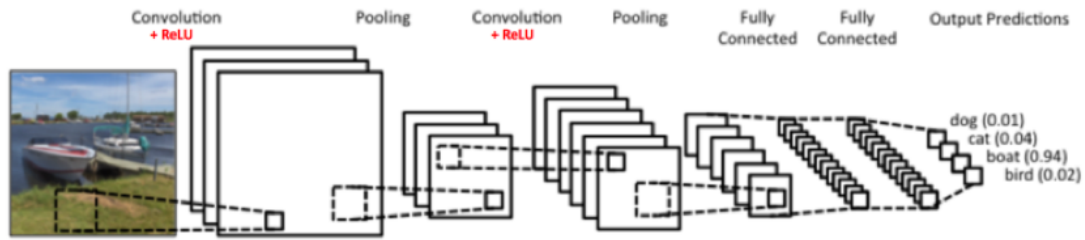


Figure B.2: The architecture of a CNN including the different types of layers.

source: <https://www.clarifai.com/technology>

levels of the network.

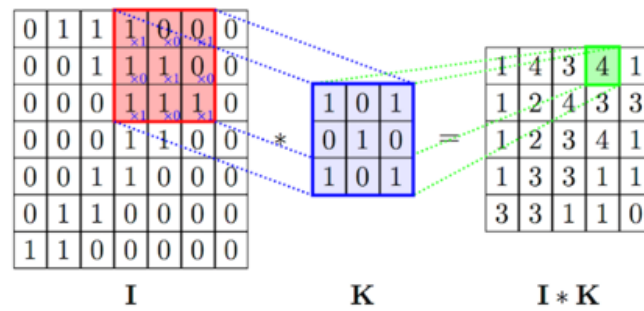


Figure B.3: The convolution operation. source:

<https://appsilondatascience.com/blog/rstats/2018/01/16/keras.html>

The ReLU is a nonlinear function that applied to the feature map; it mainly replaces negative pixels with a zero, and applies a linear transformation for positive pixels. The purpose of ReLU is to introduce non-linearity in the network. The pooling layer works as a down-sampling step; it diminishes the dimensionality of its input by passing an $(n \times n)$ object in order to extract the most important information within this object. This object can apply many functions, such as *max*, *sum* and *min*, depending on the target. An example of max pooling is shown in figure B.4.

Applying sequences of convolution, ReLU and pooling layers proved to produce high-level features for the input image. Finally, the purpose of the fully-connected (FC) layer is to use these extracted features to classify the input image,

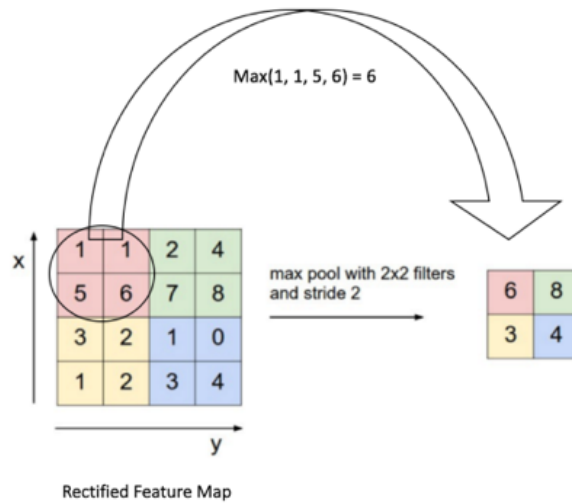


Figure B.4: The pooling operation. source:

<https://appsilondatascience.com/blog/rstats/2018/01/16/keras.html>

as shown in Figure B.2.

B.3 Entity Embedding

It's a way to represent categorical variables in a new variable space. It is mostly used in natural language processing applications. Recently, it is being used to represent any sort of categorical variables.

This approach is used as a replacement for the one hot encoding technique. Categories with various unique features, we can get sparse data on a large scale depending on the number of categories we have. Additionally, each vector obtained from the one hot approach is equidistant from other vectors. This causes us to lose information of relationships between variables. Representing categories in embeddings are a solution to dealing with categorical variables while simultaneously avoiding a lot of the downsides of one hot encoding.

An Embedding layer is a Neural Network layer that assembles categorical values holding the same label into an N-dimensional space. This representation

allows us to obtain inherent properties from each categorical value. This output can be later on used as a replacement to the one hot encoding method.

In Figure B.5 we display the model structure as interpreted from the original paper. for every category feature there's a specified input layer which then are then converted to embeddings throughout the network structure in a manner similar to word2vec.

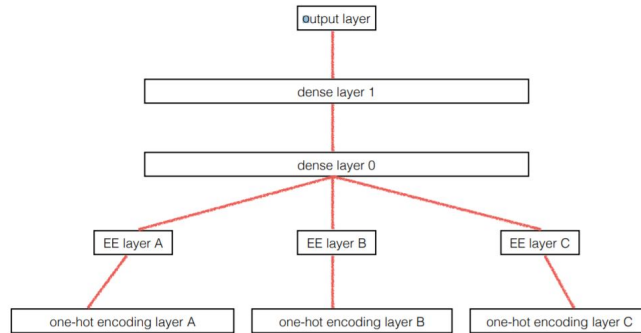


Figure B.5: Entity Embedding operation [2]

Bibliography

- [1] “DermNet skin disease atlas.” `dermnet.com`. Accessed: 2020-09-10.
- [2] C. Guo and F. Berkhahn, “Entity embeddings of categorical variables,” *arXiv preprint arXiv:1604.06737*, 2016.
- [3] S. Latif, R. Rana, J. Qadir, M. Imran, S. Younis, *et al.*, “Mobile health in the developing world: Review of literature and lessons from a case study,” *IEEE Access*, 2017.
- [4] A. Ohtani, T. Suzuki, H. Takeuchi, and H. Uchida, “Language barriers and access to psychiatric care: a systematic review,” *Psychiatric Services*, vol. 66, no. 8, pp. 798–805, 2015.
- [5] J. G. Kahn, J. S. Yang, and J. S. Kahn, ““mobile’health needs and opportunities in developing countries,” *Health Affairs*, vol. 29, no. 2, pp. 252–258, 2010.
- [6] S. Kumar, W. J. Nilsen, A. Abernethy, A. Atienza, K. Patrick, M. Pavel, W. T. Riley, A. Shar, B. Spring, D. Spruijt-Metz, *et al.*, “Mobile health technology evaluation: the mhealth evidence workshop,” *American journal of preventive medicine*, vol. 45, no. 2, pp. 228–236, 2013.
- [7] D. H. Peters, A. Garg, G. Bloom, D. G. Walker, W. R. Brieger, and M. Hafizur Rahman, “Poverty and access to health care in developing coun-

- tries,” *Annals of the New York Academy of Sciences*, vol. 1136, no. 1, pp. 161–171, 2008.
- [8] D. M. O’Connor, O. S. Jew, M. J. Perman, L. A. Castelo-Soccio, F. K. Winston, and P. J. McMahon, “Diagnostic accuracy of pediatric teledermatology using parent-submitted photographs: a randomized clinical trial,” *JAMA dermatology*, vol. 153, no. 12, pp. 1243–1248, 2017.
- [9] Y. LeCun, Y. Bengio, *et al.*, “Convolutional networks for images, speech, and time series,” *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
- [10] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144–152, 1992.
- [11] M. Portela and C. Granell-Canut, “A new friend in our smartphone? observing interactions with chatbots in the search of emotional engagement,” 2017.
- [12] E. Riloff and M. Thelen, “A rule-based question answering system for reading comprehension tests,” in *Proceedings of the 2000 ANLP/NAACL Workshop on Reading comprehension tests as evaluation for computer-based language understanding systems-Volume 6*, pp. 13–19, Association for Computational Linguistics, 2000.
- [13] C. Griffiths, J. Barker, T. Bleiker, R. Chalmers, and D. Creamer, *Rook’s Textbook of Dermatology, 4 Volume Set*. John Wiley & Sons, 2016.
- [14] L. Noueihed and D. Khraiche, “Lebanon’s economic crisis is spinning out of control, fast,” Jul 2020.
- [15] “Coronavirus disease (covid-19).” <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed: 2020-09-10.

- [16] “Beirut port explosion.” <https://www.bbc.com/news/topics/c88p951myv0t/beirut-port-explosion>. Accessed: 2020-09-10.
- [17] K. Falconer, *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.
- [18] N. Codella, J. Cai, M. Abedini, R. Garnavi, A. Halpern, and J. R. Smith, “Deep learning, sparse coding, and svm for melanoma recognition in dermoscopy images,” in *International Workshop on Machine Learning in Medical Imaging*, pp. 118–126, Springer, 2015.
- [19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 675–678, ACM, 2014.
- [20] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [21] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, “Automated melanoma recognition in dermoscopy images via very deep residual networks,” *IEEE transactions on medical imaging*, vol. 36, no. 4, pp. 994–1004, 2017.
- [22] A. Menegola, M. Fornaciali, R. Pires, S. Avila, and E. Valle, “Towards automated melanoma screening: Exploring transfer learning schemes,” *arXiv preprint arXiv:1609.01228*, 2016.
- [23] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [24] S. Demyanov, R. Chakravorty, M. Abedini, A. Halpern, and R. Garnavi, “Classification of dermoscopy patterns using deep convolutional neural networks,” in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pp. 364–368, IEEE, 2016.

- [25] J. Kawahara, A. BenTaieb, and G. Hamarneh, “Deep features to classify skin lesions,” in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pp. 1397–1400, IEEE, 2016.
- [26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [27] J. Kawahara and G. Hamarneh, “Multi-resolution-tract cnn with hybrid pretrained and skin-lesion trained layers,” in *International Workshop on Machine Learning in Medical Imaging*, pp. 164–171, Springer, 2016.
- [28] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.
- [30] H. Liao, “A deep learning approach to universal skin disease classification,” *University of Rochester Department of Computer Science, CSC*, 2016.
- [31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.
- [32] F. Chollet, “Xception: Deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.
- [33] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” *arXiv preprint arXiv:1602.07261*, 2016.

- [34] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, “Self-training with noisy student improves imagenet classification,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10687–10698, 2020.
- [35] A. Bourouis, A. Zerdazi, M. Feham, and A. Bouchachia, “M-health: skin disease analysis system using smartphone’s camera,” *Procedia Computer Science*, vol. 19, pp. 1116–1120, 2013.
- [36] L. De Greef, M. Goel, M. J. Seo, E. C. Larson, J. W. Stout, J. A. Taylor, and S. N. Patel, “Bilicam: using mobile phones to monitor newborn jaundice,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 331–342, ACM, 2014.
- [37] D. M. West, “Improving health care through mobile medical devices and sensors,” *Brookings Institution Policy Report*, vol. 10, pp. 1–13, 2013.
- [38] M. G. Dixon, I. J. Schafer, *et al.*, “Ebola viral disease outbreak—west africa, 2014,” *MMWR Morb Mortal Wkly Rep*, vol. 63, no. 25, pp. 548–51, 2014.
- [39] “Wireless Heart Health mobile-enabled rapid cardiovascular screening improves health care for rural patients in china.” <https://www.qualcomm.com/media/documents/files/china-heart-health.pdf>.
- [40] D. Coalition, “Rare diseases clinical research network (rdcrn) publications,”
- [41] J. Barth, J. Klucken, P. Kugler, T. Kammerer, R. Steidl, J. Winkler, J. Hornegger, and B. Eskofier, “Biometric and mobile gait analysis for early diagnosis and therapy monitoring in parkinson’s disease,” in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 868–871, IEEE, 2011.

- [42] M. Matthews, S. Abdullah, G. Gay, and T. Choudhury, “Tracking mental well-being: Balancing rich sensing and patient needs,” *Computer*, vol. 47, no. 4, pp. 36–43, 2014.
- [43] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, “A survey of mobile phone sensing,” *IEEE Communications Magazine*, vol. 48, no. 9, 2010.
- [44] C. N. Dos Santos and M. Gatti, “Deep convolutional neural networks for sentiment analysis of short texts.,” in *COLING*, pp. 69–78, 2014.