Connect - Doctor at Home

Abhya Singh ¹, Ayush Dubey ², Saurabh Jaiswal ³, Siddharth Singh ⁴, Prakash Joshi ⁵

^{1, 2, 3, 4} Student, ⁵ Assistant Professor

Department of Computer Science & Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, Uttar Pradesh, India.



Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 3 (May-June 2023)

License: Creative Commons Attribution-ShareAlike 4.0 International License



Abstract

In the technologically advanced world where mobile hardware is getting sophisticated as the day passes by, the performance index of the software configuration is much higher than the actual requirements. A smartphone's features depend much more on the software than the hardware and Android as well as IOS are for sure leading the smartphone world, and applications based on the Android and iOS SDK appeals to a much larger crowd. Connect - Doctor at Home will take care of you and will give you the best prescription as well as fresh information about home treatments based on the symptoms you are experiencing using CNN. It will give a fast and accurate detection and will help in maintaining the health chart of the society. This way we can get them cautious about the situation and then they can take care of their health.

Keywords: Disease Detection, CNN, Machine Learning, Deep Learning

1. Introduction

A smartphone is acknowledged as a significant technical development that has a range of effects on human lives. As the most popular operating system for smartphones, Android surpasses iOS. Anyone can use Android because it is a free and open-source operating system.

These days, technology is a part of every facet of human life. Because of their rising popularity, cell phones have attracted the attention of practically everyone. Users can send and receive messages, access the internet, consume content, and record audio and video in addition to making and receiving calls. Inbuilt features on smart phones include a touch screen, high-resolution camera, front-facing camera for video conferencing, and a keypad. Many different operating systems are supported by smartphones. Small pieces of software known as apps can do almost anything and can be downloaded or installed. With multi-core processors, terabytes of RAM, operating system, apps transform smartphones into portable computers. Initially, informational uses were allowed for smartphone apps like Gmail, calendar, and weather reports. As technology and user expectations advanced, developers started to produce apps for new purposes, including games, banking, video chats, and other activities. An app can show data in a manner comparable to a website, with the added advantage of downloading content that can be utilised offline in the event that the Internet is down. Today's market offers a wide range of programmes for

several operating systems, including Android, iOS, and others, with Android currently holding the largest market share The higher availability of electronic health record systems has given way to tremendous promises in development of many high risk disease prediction models that helps in prediction of different harmful diseases. As there are many hidden patterns available in the medical database, there is a greats requirement for studying the prediction models which further help in reducing hospital admission and medical cost. Machine learning methods are used for prediction in several domains as the results obtained are more accurate and adequate better results. Various machine learning techniques used for detection of high risk diseases are studied and conferred in this paper.

Machine learning has advanced significantly in comparison to other approaches to deciphering difficult data, and its deep learning component has emerged as the most significant subfield. Deep learning outperforms prior technical approaches in terms of its capacity to analyze and evaluate complicated data, its method for identifying the key characteristics of multidimensional data, its effectiveness in handling unstructured data, and its more accurate classification strategy. Additionally, it makes it possible for more people to learn, comprehend, and dive into the world of deep learning. The advancement of machine and deep learning technology has accelerated over the last few years. It has brought about sufficient changes in several sectors, including Speech recognition, natural language processing, and image recognition. Meanwhile, the use of deep learning for disease prediction has quickly gained popularity and produced a number of excellent findings.

The artificial neural network (ANN), also referred to as a perceptron and made up of numerous neurons organized according to a specific level, is the basis of deep learning. The neural network has successfully grown from a shallow model to a deep model, and structurally becomes a multilayer perceptron, also known as a deep neural network (DNN), with continual change in demand for models, the expansion of data scale, and the development of computing capacity.

A deep neural network consists of three components: an input layer, a hidden layer, and an output layer. The standard multi-layer perceptron neural network relies on backpropagation during training.

It entails both the forward transmission of information and the back propagation method for loss function optimisation. We may compare it to having a computer create a number of programmes that are like to the human brain and train it to think like a person so that it can mimic the workings of human eyesight, hearing, and other cognitive behaviours and improve over time. In a similar way, while creating disease predictions, a computer can be used to simulate medical professionals in order to make diagnoses and compile prior knowledge through a process known as continuous practise to increase the model's accuracy and sturdiness.



Figure 1: Deep Neural Network

Convolutional Neural Network (CNN)

The number of parameters grows as a neural network with only a fully connected layer advances, increasing the memory footprint and the computational cost. Eventually, a technique was looked into in order to resolve the issues brought about by entirely linked layers and to accelerate neural network growth to a greater extent. Utilising the ideas of local correlation and weight sharing, a convolutional neural network (CNN) is built, which not only reduces the number of parameters but also increases the effectiveness of training. LeNet's proposal, which outlined the fundamental principles of CNN, gave birth to the organisation in 1998. The basic structure, which consists of a convolutional layer, a pooling layer, and a fully linked layer, is depicted in the picture below.



Recurrent Neural Network (RNN)

Models must learn features from a sequence in order to be used in applications such as task-based discourse and time series forecasting. The failure to correctly regulate the time series results from the neurons at each layer of the neural network, including the fully connected layer, being unable to link to one another. CNN has some spatial problem-solving capabilities, but not in the area of time.

Neurons and feedback loops make up RNNs. In situations when the first and second inputs are interdependent, RNN has clear advantages. When processing massive amounts of delayed and erratic

noise, RNN performs better than the majority of machine learning algorithms. However, during training, RNN runs the danger of gradient explosion and disappearance.



Logistic Regression is a method used for classify which is an advanced way for ordinary regression. This model is capable of modeling only a forked variable that can be implemented on the occurrence or nonoccurrence of an event. This method is helpful for analysing the probability of the condition on which a new instance may belong to a specific cluster of class. The ability to predict chronic diseases has been increasingly popular among researchers as it has created numerous opportunities for new study and development.

Naïve Bayes is a classification technique used by the research community which is mainly used for training and classification and this algorithm is basically derived from the Bayes' theorem. This theorem is applied in finding the probability of an event from the knowledge base that is derived from the training model which is related to the condition of that event. This classifier derives its idea based on the assumption of a particular feature in a class not related to any other feature of the same class, even though features belonging to that class are interdependent on themselves.

In machine learning, several algorithms are there which belong to different families. However, all algorithms are not suitable to detect all types of diseases. So, providing the models for the detection of disease based on the symptoms is also a matter of challenge. As a matter of fact, these models are used to predict the disease in one direction by applying a uniform logic. For example, the decision tree applied the tree-based method; the linear model applied the sigmoid rule etc. So, there may be lagging in the identification process as not all data may be suitable for the unified process. Thus, the application of hybrid models plays an important role by accumulating heterogenous state-of-the-art models so that the detection of the disease can be done perfectly. In this present research, the hybrid model will be applied to propose to detect different kinds of diseases.

2. Literature Review

The application of machine learning algorithms is at peak because of its efficiency to create highly capable learning models. It uses many types of statistical and optimization methods to create training modules and to derive patterns from the past data and use the same of prediction effectively. Popular algorithms of machine learning are used in various areas like automated text categorization, network intrusion detection, junk e-mail filtering, optimizing manufacturing process and disease modeling and in a lot of different areas. Off late supervised algorithms are more preferred than unsupervised algorithms because of various challenges. In a supervised learning model, a training model is built based on training the dataset multiple times. As the data is labeled, it becomes easy to predict the outcome of unlabeled data. Recent studies have found that AI and ML techniques play a very important part for the prediction of high risk diseases. Machine learning methods can be applied to individual patient data which may fall under different research directions. The research finds are very useful as they help health professionals to create awareness among patients to prevent such high risk diseases.



Heart Disease Detection

Applied machine learning models to detect heart disease. To execute the research, they selected the heart disease dataset from the UCI repository. They have reviewed various existing methods and approaches and have detected that the reasons for heart disease may vary by a person according to the symptoms. This is one of the important reasons why doctors use to prescribe different medicines for the same heart disease to different patients because of the change in the symptoms. So, they have emphasised symptombased detection. They used different machine learning models and finally observed that mostly the voting process detected heart disease with the highest accuracy with an accuracy rate of 90%. So, finally, they have proposed the model for the detection of heart disease with the majority voting ensemble model. Kumari & Mehta (2021) [7] have applied machine learning to detect heart disease. While observing the efficiency of the machine learning model for spotting the disease, especially heart diseases, they have observed that not all machine learning models work well for this purpose. In some of the cases, they have observed very low accuracy in detection by the models of machine learning which are generally known as weak learners. So, they have proposed a new model for the detection of heart disease by improving the efficiency of the weak learner with the application of Adaptive Boosting and Voting ensemble models. With the application of ensemble models, they have obtained 84.7% accuracy in detecting heart disease with no or very less model over fitting.

Chaurasia and Pal have recommended using data mining techniques to find cardiac disease. For mining, a collection of machine learning techniques are utilized in the WEKA data mining tool. For this viewpoint, Naive Bayes, J48, and bagging are employed. A data collection on heart disease from the UCI machine learning lab has 76 features. Only 11 criteria are used in the prediction process. Naive Bayes offers an accuracy of 82.31%. J48 provides 84.35% accuracy. The accuracy of Bagging is 85.03%. On this data set, bagging offers a higher classification rate.

To predict the risk of diseases using both structured and unstructured patient data, we propose the use of convolutional neural network techniques. Our model achieves an accuracy range of 85% to 88%. Additionally, we suggest employing different machine learning algorithms on the training data to determine the most effective one for predicting disease risk.





Cancer

Any Machine learning algorithm requires preprocessing, feature selection or extraction and then classification. Diagnosis of cancer depends on Feature extraction, which can be extended to detect cancer setting to benign and malignant tumours. MS Uzer et al. has developed a breast cancer detection model which uses the concept of Artificial Neural Networks. The classification mode used is a combination of Backward and Forward Sequential Searching with PCA, using which an accuracy of 98.75% was achieved.

Author presented usage of a neural wavelet network which looks at a linear and localised approach to detect cancer. The algorithm is trained on its parameters using a Recursive Least Squares (RLS) technique. The findings when compared to previous trials, demonstrated that the suggested technique is highly efficient and offers an excellent categorization To increase the accuracy of CAD-based procedures, boosting has been used by the author. A hybrid boosting approach was developed, to approach the problem of characterization of breast cancer, this combines the benefits of multiple boosting techniques. When evaluating different boosting techniques on real breast cancer data, the hybrid boosting algorithm demonstrated superior performance, surpassing the other boosting methods by 48%.

For automating breast cancer diagnosis by combining a classifier based on Support Vector (SVM) networks with a neural network (ANN), an intelligent system was created. Wisconsin diagnosis records were used to train the SVM on the dataset, which helped to discriminate between benign and malignant clusters.

Cancer has a protracted development cycle and the potential for many gene alterations. To gain more effective knowledge, more data and associated genes must be evaluated. Because it can obtain a more full flow of sickness information and significantly improve the accurate identification of cancer disorders, the multi-omics joint technique will emerge as a new trend in the field of cancer research. There are numerous tumor data sets in the UCSC Xena database. There are many multi-omics data sets available through the Cancer Genome Atlas (TCGA). It encourages the study of cancer diseases using several omics. Tan et al. proposed the MOSAE (Multi-omics Supervised Autoencoder) model for genomics research, comparing it with six other models and employing a specific AE to achieve the processing and fusing of various omics structures.

The dataset utilized in these experiments include Fine Needle Aspirates-based measures (FNA). Most of the work comparing various traditional statistics-based approaches along with traditional Machine Learning (ML) classification procedures has been published in order to highlight the benefits of (ML) and its potential.

When processing data on diseased cells, the DL model can more effectively pinpoint potential characteristics. It can then use AE or GAN to enhance the data, solve unlabeled and irregular data, and optimise noise data to detect cancer cell disease. For the diagnosis and classification of cervical cancer, Khamparia et al. used CNN and variational autoencoder (AE). AE reduces dimensionality while enriching the input, followed by CNN to extract high-dimensional features, and then the data is delivered to the dense layer and Softmax layer for classification. When the size is 3 3, an accuracy of 99.4% is reached after using different filters and comparing to the current approach; this is better than the comparison method. The InceptionV3 model is employed by Iizuka and colleagues.

New cancer treatment approaches will arise as a result of technological developments in medicine and the introduction of a new generation of computers. Radiologists manually carry out conventional cancer diagnosis techniques. A medical scan can have millions of pixels, but the malignant area may only take up a few dozen. Manual diagnosis takes time and contains errors. Furthermore, the requirements for radiologists are more stringent. In order to provide a more thorough diagnosis base, deep learning can be utilised to identify deeper cancer cell lesions while also taking horizontal and vertical features into account. The strain on medical diagnosis can be alleviated to a greater extent, and it can improve the shortcomings of artificial care while achieving earlier and more precise cancer cell diagnosis.

Takahashi and colleagues utilized autoencoders to analyze six multi-omics datasets, reduce dimensionality, and unify features in order to create a lung cancer prognosis prediction model using machine learning. The model exhibited several advantageous indicators. Despite conducting multiple keyword searches across various omics, there is still a relatively limited amount of research on using deep learning and multiple omics for predicting cancer diagnosis. The advancement of genomics technology has made it easier to gather cancer genome data. Going forward, we should fully leverage the benefits of DL technology for multi-data processing and focus on integrating multi-omics research. This

will allow for a more comprehensive identification of cancer cell status from various perspectives, given the dynamic and unpredictable nature of cancer development.

Kumari and Chitra conducted an experiment to predict diabetes illness. SVM is the machine learning technology utilised by the scientist in this experiment. SVM employs the RBF kernel for classification purposes. A machine learning laboratory at the University of California, Irvine provides the Pima Indian diabetes data set. To carry out the experiments, MATLAB 2010a is employed. SVM has a 78% accuracy rate.





Liver Disease

Alfisahrin & Mantoro (2016) have used the classifiers of machine learning to detect liver disease. To conduct the research, they selected the dataset known as the Indian Liver Disease dataset from which they have selected the important features. Based on the selected features, the researchers have prepared and finalized the data. They have applied the classifiers such as Naïve Bayes, Decision Tree etc. to classify the symptoms and detect liver disease. Finally, they predicted liver disease with the application of a Decision Tree with 92% accuracy. Sivasangari et al. (2020) have emphasized the detection of liver disease by detecting important symptoms. In this context, they have used the feature selection process through correlation and obtained the final features. The study's goal is to use medical image analysis to

solve medical problems by studying diagnostic images gathered in clinical settings. The major purpose is to obtain meaningful and efficient data for better diagnosis. Due to their superior performance, researchers in the field of medical image analysis have increasingly turned to deep learning, specifically CNNs.

Brain Tumor

Uncontrolled and rapid cell proliferation in the brain is the primary cause of tumor formation, which if left untreated, can be fatal. Despite many significant efforts and promising outcomes, accurate segmentation and classification of brain tumors remain challenging tasks. Variations in tumor location, size, and shape pose significant challenges for the identification of brain tumors. This survey provides a comprehensive literature review of the use of magnetic resonance imaging (MRI) in detecting brain tumors, including the tumor's structure, publicly available datasets, augmentation techniques, segmentation, feature extraction, classification, deep learning, transfer learning, and quantum machine learning. The study also includes an assessment of relevant research on brain tumor detection. Specifically, the investigation considers the architecture of brain tumors, publicly available datasets, augmentation methods, segmentation, feature extraction, classification, classification, deep learning, transfer learning, transfer learning, and quantum machine learning. Additionally, the study evaluates all significant research on brain tumor detection. The study employed an LSTM model to address the classification issue.

This approach smooths input MRI images using N4ITK and a 55 Gaussian filter before passing them as input to a four LSTM model. The four hidden Units 200, 225, and 200, are used to build the LSTM model.

This model's effectiveness has been evaluated using the BRAT (2012–2015) series and the SISS–2015 benchmark dataset. This study combines various MRI sequence types to create a brand-new framework. Comparing the fused sequence to a single sequence reveals more information. Later, the 23–24 CNN models received a fused sequence. The suggested model is trained to detect gliomas using data from the Brat series. For the purpose of detecting glioma and stroke lesions, the 14-layer CNN model has been trained entirely from scratch using six Brats series datasets. ELM and RELM classifiers are used to perform the classification.

The principles of entanglement, parallelism, and quantum state superposition can be leveraged to create powerful quantum computers. However, the lack of computational resources for running quantum algorithms poses a challenge to exploring the potential of quantum properties for efficient computation. With the development of quantum technology, conventional computers based on quantum theory and impacted by qubits struggle to fully exploit the benefits of quantum state and entanglement. Quantum Artificial Neural Networks (QANN) have demonstrated usefulness in various computer tasks such as classification and pattern recognition due to the inherent properties offered by quantum physics. Nevertheless, quantum models based on real quantum computers use large quantum/qubits as a direct representation of matrices and linear functions.

By using atlas-based registration, Rupsa Bhattacharjee's technique offers the potential to implicitly partition brain pictures containing tumors. detection of anomalous mass accumulation has a significant impact on the image's pixel-wise intensity distribution. The tumour mass and volume are present in the processed output exactly as they were in the original MRI image. The approach proposed by Mircea Gurbin makes use of various wavelet levels; the high precision component is achieved via CWT. They

were accurately identified as benign tumours, malignant tumours, or healthy brains using SVM. There are substantial computational benefits to using Support Vector Machines. The data in this research demonstrates that, as compared to DWT, CWT offers higher computation. In this methodology proposed by Bhavana Ghotekar, a method to detect affected brain tissues due grade 4 GBM was proposed. Naive Bayes Classification is done on the extracted properties. In this method an accuracy of 83.33% was obtained. The algorithm first takes the Magnetic resonance imaging image as input to which filtering is done to smoothen the image and remove noise from the image. The next step is masking the filtered image which is removing brain tissues from the skull. In the last stage the obtained features are given to the Support Vector Machine classifier which classifies the image as cancerous or not.

A method based on neural convolution was proposed by Fatemeh Derikvand. This method uses Glioma Brain Tumor Segmentation Networks in Magnetic resonance imaging and the process proposed is a mixture of Multiple different Convolution Neural Network architectures which makes use of Local and global knowledge of brain tissue to predict the label of each pixel, which helps to enhance results. In another paper proposed by G. Hemanth Convolution Neural Network has been employed for Segmentation of images. It explicitly extracts features from Images of pixels with the minimum preprocessing involved. The network used is LinkNet, a light, deep network. Architecture of a neural network built to perform Segmentation of Semantic. Contains the LinkNet Network Blocks of encoders and decoders which basically manage to Break the picture and rebuild it again until it is sent through a few final layers of convolution. The proposed Root Mean Square Error (RMSE), recall, sensitivity, precision, F-score specificity, and PME (probability of the misclassification error) are used to measure the performance of convolutional neural networks. In the methodology proposed by Tonmoy Hossain firstly skull striping is done to the input image which removes the skull portion from the Magnetic resonance imaging image. This filtered image segmentation is done using Fuzzy C-Means clustering algorithm. For Segmentation we only need the brain part rather than the skull for this we do morphological operations in our images.

COVID Detection

The use of deep learning and attention mechanisms in graphical-based symptomatic detection techniques for COVID-19 diagnosis has shown promise in improving the accuracy and efficiency of screening and diagnosing patients. By training deep learning models on large datasets of CXR and CT scan images, patterns and cluster-like structures associated with COVID-19 can be identified and used to develop computer-aided diagnostic (CAD) technology.

The technique proposed by Shah et al. (2021) involves a knowledge distillation network topology, where a teacher network gathers material from worldwide sources and emphasizes the model, while a student network receives attention and is responsible for making predictions on new data. The attention mechanism used in this technique improves the ability of the model to differentiate between tasks involving processing X-ray images.

Overall, the use of deep learning and attention mechanisms in graphical-based symptomatic detection techniques shows promise in improving the accuracy and efficiency of COVID-19 diagnosis, particularly in cases where traditional radiographic imaging techniques may not be readily available or feasible. However, further research and validation are needed before these techniques can be widely adopted in clinical settings.

Haritha et al. (2020a) proposed a strategy for identifying the location of COVID-19 infection using Xray images of the pulmonary region. However, due to limited medical data and computing power, the authors suggested two approaches. The first involved creating a custom convolutional neural network (CNN) model and training it on a large dataset of non-COVID-19 chest X-ray images (ChexPert), followed by fine-tuning the model with COVID-19 X-ray images. However, this approach did not yield the best results. Alternatively, transfer learning using pre-trained CNN models such as Resnet50, VGG16, and Densenet121 was suggested. The Densenet121 CNN model achieved an accuracy of almost 90%, which was promising for identifying COVID-19 infection sites using X-ray images of the pulmonary region.



Figure 7: Performance Metrics for Binary Classification (Haritha et al., 2020a)

Haritha et al. (2020a) proposed a method for identifying the site of infection using X-ray images of the pulmonary region, but due to limited medical data and processing resources, two approaches were suggested. The first approach involved creating a custom convolutional neural network (CNN) model and training it on a large dataset of non-COVID-19 chest X-ray images (ChexPert). Fine-tuning the model with COVID-19 X-ray images did not produce the best results. The second approach involved transfer learning using pre-trained CNN models such as Resnet50, VGG16, and Densenet121. Using the Densenet121 CNN model, an accuracy of almost 90% was achieved for identifying COVID-19 infection sites using X-ray images of the pulmonary region.

Knipe (2020) developed CoviNet, a multi-layered CNN architecture, to address class imbalances in the dataset and automate feature extraction. Rectified Linear Activation Unit (ReLU) was used as the activation function after each convolution layer, and max-pooling layers were introduced to reduce computation complexity and produce feature maps. The output was then flattened and transferred to the fully connected layer, with dropout used as a regularisation technique to avoid overfitting. Future research can focus on using larger clinical datasets to further evaluate the model's efficacy and better segment the data.

In the research conducted by Huang L. et al. (2020), they proposed a novel CNN framework for the detection of COVID-19. The CNN architecture consists of 12 weighted layers that include 2 convolutional and 12 fully connected layers. The fully connected layer receives a 2D feature vector as

input and sends it to the SoftMax classifier. The output layer contains two neurons to classify the images into two categories: COVID-19 positive or negative. The dataset used in this study consisted of 625 images, which were divided into training, testing, and validation subsets. The training subset was used to construct the model, while the testing subset was used to evaluate the accuracy of the developed model. The new CovStacknet distribution model, which is based on the StackNet meta-modeling technique, is described in another article (Qjidaa et al., 2020b). The article makes use of three datasets. There are 1,553 normal cases and 4,273 pneumonia cases among the 5,216 X-ray images in the first dataset (DS1). Kaggle provided the second dataset (DS2) freely available for exploration. The University of Montreal made the third dataset available to the general public. In this work, the necessary characteristics from X-Ray were extracted using a convolutional neural network called VGG 16. The StackNet classifier will use 25,088 real array values from the feature set as features for each image.





An automated method is used in the study suggested by Rabbah et al. (2020) to find COVID-19. Images from CT scans are inputted. It employs predefined image processing algorithms as a preliminary step. These image processing techniques aid in pinpointing the precise lung area. The remainder of the image's area that doesn't convey important information is discarded. These preprocessing techniques reduce classification errors. The accuracy of recognising COVID-19 from CT Scan pictures has been greatly enhanced by the techniques suggested by the authors. This research primarily employs a feature pyramid network created for specialised classification tasks. The model will be able to study various image resolutions thanks to this preprocessing without losing information from microscopic objects.

An automated distribution segmentation mechanism was suggested by researchers (Dodds et al., 2020) to help diagnose COVID-19 chest CT scans. An innovative deep learning-based approach to detect COVID-19 infection was presented in this research. With the aid of CT scans, this work was further expanded to section COVID-19 anomalies. Segmentation, classification, and reconstruction are carried out simultaneously using various datasets. A multilayer perceptron was used to develop a common encoder with three jobs for reconstruction and segmentation and a decoder for disentangled feature representation. The suggested technique was tested on a set of 1,369 patient lung images, 449 of which had COVID-19, 425 of which are normal, 98 of which have lung cancer, and 397 of which have other disorders.

Majeed and Hwang's study from 2022 gives insight on the use of artificial intelligence (AI) and other cutting-edge technologies in the fight against COVID-19. This study provides a thorough understanding of the technology supporting early detection/diagnosis, trends analysis, intervention planning, forecasting the burden of healthcare, comorbidity analysis, mitigation, and control. On a broader scale, this study has provided the field of artificial intelligence with novel applications for successfully addressing the COVID-19 pandemic issue.

Skin Cancer

To tackle the challenges of data imbalance and scarcity in skin disease datasets, the International Skin Imaging Collaboration (ISIC) developed the HAM10000 dataset. This dataset features more than 10,000 dermoscopic images of pigmented skin lesions, covering seven representative skin diseases. These diseases include basal cell carcinoma, benign keratosis, cutaneous fibromas, melanoma, melanocytic nevi, and vascular lesions like hemangiomas, purulent granulomas, and subcutaneous hemorrhage. The Derm7pt dataset contains around 2,000 clinical and dermoscopy color images of skin disorders and structured data for testing and training CAD systems. It serves as a database for analyzing the prediction outcomes of the seven-point skin lesion malignancy checklist.

The ISIC Archive has made a publicly accessible skin disease dataset available to the computer science community worldwide to lower skin cancer mortality while advancing and applying digital skin imaging. It currently includes over 13,000 typical dermoscopic images from clinical settings globally that have undergone expert inspection and annotation to ensure image quality. Most studies using the ISIC dataset have focused on classifying and segmenting skin cancer, with binary classification being the most prevalent.

This paper proposes a Convolutional Neural Networks-based method for classifying melanoma to assist patients and medical professionals in distinguishing benign and malignant skin cancer groups. According to the experimental and assessment section, the model can serve as a baseline for aiding medical professionals in identifying skin cancer.

3. Proposed Methodology

The software for the suggested system is created using Flutter, which is quickly becoming the most popular operating system for smartphones and tablets. Because of its increased open source programming flexibility for third parties, Flutter — which is based on all operating systems having a user interface — was chosen for this job. Therefore, using this platform for development does not require a license or any other kind of payment. Both the Android and iOS applications' code is created using the Dart programming language. The generated APK files are uploaded to the smartphones' device

managers after being downloaded to them through Bluetooth or a high-speed micro USB data transmission cord. corresponding application with a distinct user ID. The installed app will display on the phone's screen with the specified user ID.

The primary objective is to end the present practice of requiring patients to visit a doctor in order to have their illnesses identified before receiving treatment. After collecting a large dataset of medical images and relevant patient information, pre-process the images by resizing and normalising them to ensure consistency in the input for the model. Then an appropriate CNN architecture such as VGG, ResNet, or Inception and design the model by selecting the number of layers, filters, and other hyperparameters. After that use the pre-processed dataset to train the model. Monitor the performance of the model using metrics such as accuracy, precision, and recall. Use the trained model and integrate it into the app using the Flutter framework. After integration of the model, test the app using a separate dataset and evaluate its performance using the same metrics used during the training phase. After testing and making any necessary adjustments, deploy the app to the app stores or make it available for download.

Various modules of the app are as following:

- User Login Module: To log in to this module, a user must input his email address. The login module of an application through an email address with OTP (One Time Password) involves an extra layer of security to ensure the authenticity of the user.
- User Verify Module: In this module, the user must verify himself by clicking the link delivered to the email address he provided. A user can obtain another link by clicking obtain verify link. The user will be returned to the home screen after verification.
- User Forgot Password Module: In this module, the user has to enter his email id to get his reset password link on the email, a successful reset user can log in using his new credentials in the login module.
- User Home Screen Module: The user can view all of the health statuses for that day in this module. If the partner wants to read the prior healthcare report, the user can navigate to different modules via the menu bar.
- Search Module: In this module, the user can view the previous entries of a particular healthcare status using the date and also can save this data in a pdf.
- Edit Profile Module: In this module, a partner can modify his profile credentials such as phone number, password, and so on, except for the email address that was entered at registration, and the data will be changed.

Figure 9: Flowchart of Prediction Process



4. Result

The images were trained for 12 epochs with batch size 32 for brain tumor and lung cancer disease. The model achieved the training accuracy of 96.33% and 91% respectively.

In case of COVID detection, images were trained for 50 epoch with batch size 256, and the model achieved the accuracy of 100%.

For skin disease detection, images were trained for 20 epoch with batch size of 128, and achieved accuracy of 97.43%.

IJIRMPS

5. Conclusion

This research work presents multiple disease predictions using histopathological images. A convolutional neural network (CNN) was implemented to classify an image of different categories. The model was able to achieve an average of 95% accuracy for different diseases.

6. Future Scope

The future scope of this project is that we can add more disease types in the application. The concept of prescription on the basis of disease can also be added in the application.

The future scope is wide and various technologies can make this application more efficient to use.

Acknowledgement

We are thankful to RKGIT, Ghaziabad for supporting the research and providing the required research platform to authors.

References

- Yinhao Wu, Bin Chen, An Zeng Dan Pan, Ruixuan Wang, Shen Zhao, "Skin Cancer Classification With Deep Learning: A Systematic Review", Frontiers in Oncology, July 2022 Volume 12, Article 893972.
- 2. Md. Fahimuzzman Sohan, Anas Basalamah, Md. Solaiman, "COVID-19 Detection using Machine Learning : A large scale assessment of x-ray and CT image datasets", Journal of Electronic Imaging, March 2022.
- 3. Nareen O.M. Salim, Mohsin Abdulazeez, "Human Diseases Detection Based On Machine Learning Algorithms : A Review", International Journal of Science and Business, Volume 5, Issue 2, Pages 102-113, 2021.
- 4. Kavita Saxena, Rishabh Sharma, Rishav Kumar, Roshan Kumar, "Disease Prediction Using Machine Learning and Deep Learning", IJRASET, Volume 10, Issue V, May 2022.
- 5. P. Gokila Brindha, M. Kavinraj, P. Manivasakam, P. Prasanth, "Brain tumor detection from MRI images using deep learning techniques", IOP Conference Series: Materials Science and Engineering, 1055, 2021, 012115.
- 6. Bijaya Kumar Hatuwal, Himal Chand Thapa, "Lung Cancer Detection Using Convolution Neural Network on Histopathological Images", International Journal of Computer Trends and Technology, Volume 68, Issue 10, Pages 21-24, October 2020.
- Akanksha Kumari, Ashok Kumar Mehta, "A Novel Approach for Prediction of Heart Disease using Machine Learning Algorithms", 2021 Asian Conference on Innovation in Technology (ASIANCON), Pune, India, August 2021, <u>https://doi.org/10.1109/ASIANCON51346.2021.9544544</u>
- 8. https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000
- 9. https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia
- 10. https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database
- 11. https://github.com/sartajbhuvaji/brain-tumor-classification-dataset