© June 2025 | IJIRT | Volume 12 Issue 1 | ISSN: 2349-6002

AI - Driven Disease Detection Using Machine Learning and Deep Learning

Chhoti Kumari[1], Anoop Kumar Sahani[2], Archit Siwach[3], Mr. Ritesh Kumar Tripathi[4]
[1,2,3]Dept. computer science & Engineering(IOT) Noida Institute of Engineering and Technology Greater Noida, India
[4]Assistant Professor, Noida Institute of Engineering and Technology Greater Noida, India

***Abstract*—Increased outbreaks of conditions like diabetes and dermatological diseases have made effective disease diagnosis systems an essential tool to prevent such occurrences early enough. This paper outlines the development of a web system for predicting disease using machine learning algorithms. It offers the client an interface platform where he can scan for likely skin conditions based on uploading of images and project diabetes through evaluating chosen symptoms. It utilizes CNNs for diagnosing dermatological conditions and Random Forest to predict diabetes risk, showing the potential of AI-based preliminary diagnostics. In addition, the system incorporates a friendly interface to improve accessibility and usability. It also has an easy-to-use dashboard for users to monitor their health status in the long term, enabling them to view trends and patterns in their conditions. The system can be extended to forecast other diseases by adding other datasets and machine learning models. Our findings prove encouraging accuracy, which indicates the potential of disease prediction systems driven by AI in supporting early diagnosis, lowering costs of diagnosis, and enhancing outcomes for patients.**

***Index Terms*—Disease Prediction, Machine Learning, CNN, Random Forest, Diabetes, Skin Diseases, Web Application, AI- Driven Diagnosis.**

I. INTRODUCTION

A. Background to the study

Grossly neglected health conditions like diabetes and skin diseases tend to develop severe complications. With the emergence of artificial intelligence (AI) and machine learning (ML), the development of automated disease prediction models has become a feasible solution towards helping in the early diagnosis of diseases. In this paper, a web-based disease prediction system combining CNN for skin disease detection and Random Forest for diabetes prediction is suggested. In addition, the suggested system can close the gap between accessibility and affordability of healthcare by offering a preliminary diagnosis without the necessity of immediate professional action. The system is scalable and adaptable and can integrate with healthcare databases and wearable monitors to improve predictive accuracy. Use of real-time data processing makes continuous improvement and learning of the model's efficiency possible. In addition, this system can act as a supporting tool for medical practitioners, providing information on disease trends and supporting data-driven medical decisions.

B. Overview

The rising incidence of diabetes and skin conditions, which are frequently overlooked, poses the need for early detection to avoid serious complications. The recent developments in artificial intelligence (AI) and machine learning (ML) have also made it possible to create computerized disease prediction models, which can serve as an affordable and accessible option for early diagnosis. The work describes a web-based disease predictive system that applies Convolutional Neural Networks (CNN) to detect skin disease and Random Forest to predict diabetes. Using these AI methods, the system can evaluate user-supplied skin images and symptom-based data to create initial diagnostic feedback. One of the main benefits of this system is that it fills the gap between the accessibility and affordability of healthcare by providing an initial diagnosis without necessarily calling for immediate professional intervention. The system is scalable and can be adaptable, with the possibility of being integrated into healthcare databases and wearable health monitoring technology to improve predictive power. Moreover, real-time processing of data allows for constant learning, allowing the model to improve its predictions as time passes. In addition to serving individual patients, the system also serves as a decision-support system for physicians, offering insights into disease patterns and facilitating data-driven medical decisions. With additional development, this AI-based method could transform early disease detection, making healthcare more efficient and accessible to all.

C. Problem Statement

Diabetes and skin disorders are two of the most under- prioritized health conditions, which if not detected early, cause serious complications. The lack of access to healthcare centers, high cost of diagnosis, and unavailability of specialized medical experts are some of the additional factors that add to delayed diagnosis and treatment. Conventional diagnostic techniques need expert hands and are time-consuming, and hence early intervention becomes impossible for a majority of the population. With the development in artificial intelligence (AI) and machine learning (ML), predictive models of diseases using automation can be a very viable solution for overcoming these hurdles. Current healthcare solutions based on AI, however, are typically not accessible, scalable, and integrated with live data sources. There is a pressing need for an effective, affordable, and easy-to-use system that is capable of rendering initial diagnoses without the need for instant professional treatment. This study seeks to create a web-based disease prediction system that combines Convolutional Neural Networks (CNN) for skin disease detection and Random Forest for diabetes prediction. The system is intended to fill the gap between affordability and accessibility in healthcare by offering users a trustworthy, AI-based preliminary diagnostic tool. It also aims to assist medical professionals by providing useful insights into disease trends, enabling data-driven medical decision-making. Through the use of real-time data processing, integration of healthcare databases, and wearable device support, the system in question seeks to improve predictive accuracy continuously while enhancing access to healthcare for more people.

D. Objectives

This article explores AI applications in preventive healthcare with the following objectives:

- Evaluate Predictive Accuracy: Assess the performance of CNN for skin disease detection and Random Forest for diabetes prediction in determining disease risks.
- Analyze Cost-Efficiency: Examine how AI-driven disease prediction can offer an affordable alternative to traditional diagnostic methods, improving early detection accessibility.
- Assess Public Health Impact: Investigate the role of AI in promoting early disease detection and its potential effect on reducing severe health complications in the population.
- Identify Barriers to Implementation: Highlight challenges such as limited dataset availability, model scalability, and user trust in adopting AI-based disease prediction systems.
- Enhance Integration with Healthcare Systems: Explore the feasibility of integrating the AI model with healthcare databases and wearable devices for improved predictive accuracy and real-time monitoring.

E. Scope and Significance

This project is aimed at AI-based preventive healthcare applications, i.e., early diagnosis of diabetes and skin dis- eases with the help of predictive models. With the use of Convolutional Neural Networks (CNN) for detecting skin diseases and Random Forest for predicting diabetes, the system will help in accurate and timely diagnosis to the users. The system is intended to bridge the gap between affordability and accessibility in healthcare and move from a reactive sick-care model to a proactive well-care model. The importance of this project is in its ability to reduce avoidable health complications via early intervention. The system takes advantage of AI- based forecasting algorithms to redefine healthcare pathways, streamline diagnostic procedures, and lower treatment costs over the long term. In addition, through the offer of personalized patient care suggestions, the platform enhances data- driven decision-making for both users and health practitioners.

II. LITERATURE REVIEW

There is an issue with the alteration in the error rate change in the database, due to a change in the data size utilized in various skin cancer tests. Thus, the absence of a standard database can cause severe issues; average error values are taken into account in most tests. Furthermore, data collection of various studies relies on each study, causing unnecessary effort and time. When the real class is manually marked and contrasted with the class predicted to determine the matric of a single parameter, the pixels are lost when the background is trimmed into a skin cancer image using Adobe

Photoshop. At this juncture, the process affects the results of all the loyal parameter groups (matriculants, relationships, and behaviors), which are deemed controversial. High reliability and low level of complexity cannot coexist, which is exemplified in the training program and is affected by conflicts between varying levels, giving rise to significant challenges. The mechanism utilized for the identification of one skin lesion does not necessarily function to identify others. Numerous varying courses and sets of tests have been applied to assess the proposed methods. Furthermore, under the training and testing parameters, various researchers are interested in various fields. This inconsistency in all the papers makes the right comparison impossible at all. Though these references in the literature have been under severe criticism, studies still utilize them to analyze the application to skin cancer and other imaging domains. The test data utilized is generally very small to enable a valid statement concerning system performance to be achieved. As much as it is important to recall, one cannot gather enough appropriate data on the web at this stage of information. This data, with high uncertainty, clearly does not satisfy the conditions of independent and uniform distribution, which is one of the basic conditions under which in-depth learning can be used effectively. For some rare and minor illnesses, only a finite number of pictures are present to train. Hitherto, numerous algorithms were found to discriminate against minorities, which may produce a large divide in the healthcare service among ”the rich” and ”the poor.” Numerous aspects of the training are needed using in-depth approaches. Additionally, while in-depth learning technique has been implemented successfully elsewhere in vocations, skin-augmented models function only for specific dedicated diseases and do not work under normal conditions. Diagnosis of dermatology is a complex process, in addition to image recognition, it should be supplemented with other techniques such as blushing, sniffing, temperature changes, and microscope. We have used pre trained VGG16 model .The model’s last three dense layers are customized according to our classification of diseases. The website is simple to use and increases user interactivity.

In Miotto et al. [18], the Indian Diabetes Pima (PID) dataset was obtained from the UCI Machine Learning Repository, originally created by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The data set contains 768 patients of female gender who are 21 years and older, with nine features employed to predict diabetes: count of pregnancy, BMI, level of insulin, age, blood pressure, skin thickness, glucose, diabetes genealogical function, and out- come classification. The ”result” feature is the goal variable, where 0 is non-diabetic and 1 is diabetes.

In this research work, machine learning algorithms and data mining techniques were employed to improve the accuracy of diabetes prediction based on user symptoms. Another dataset was collected from a Bangladeshi hospital that comprised self-reported symptoms and validated diabetes diagnoses by doctors. Since it performed well with machine learning tasks, Random Forest was implemented for classification [19].

A. Early Detection Models in Skin Disease Management

Early detection models based on artificial intelligence have improved the accuracy and speed of diagnosis of skin disease, enabling faster and more accurate diagnoses. Deep learning models, particularly Convolutional Neural Networks (CNNs), have been widely used to analyze dermatological scans and detect patterns associated with various skin diseases. Ac- cording to Miotto et al. [18], AI models such as VGG16 can identify the early warnings of skin disorders, including melanoma, vitiligo, and dermatitis, by capturing minor variations in skin texture, color, and lesions. Utilizing large dermatological datasets, AI-based systems enhance diagnostic performance with less dependency on manual diagnoses and fewer errors. These models also facilitate electronic health record (EHR) analysis, enabling doctors to track the progression of the disease and provide appropriate treatments. With improvements in AI technology, real-time data processing, telemedicine platforms, and wearable skin monitoring devices will further improve early detection abilities so that timely intervention and improved patient outcomes can be achieved [13]. Subsequent research must focus on improving model generalizability across different skin colors to make AI-based diagnosis of skin diseases more inclusive and global [18].

B.Early Detection Models in Diabetes Management

Early warning models based on AI have greatly enhanced diabetes prediction and prevention through the ability to con- duct faster and more accurate risk assessments. Random Forest machine learning algorithms compare patient information, such as polyuria, polydipsia, acute weight loss, and blood glucose levels, to predict the development of diabetes. Miotto et al. [18] indicate that AI models are able to detect subtle risk factors in patient history, enabling early medical intervention and lifestyle changes for disease prevention.

Along with traditional diagnostic methods, AI improves electronic health record analysis, enabling medical professionals to identify high-risk individuals and adopt customized treatment protocols. Real-time processing of data, coupled with wearable health monitoring equipment, facilitates real- time monitoring of blood glucose and other vital parameters, further solidifying preventive care programs. With the ongoing development of AI technology, its application in diabetes prediction, monitoring, and management will keep improving, leading to earlier diagnosis, fewer complications, and better patient outcomes [13].

C. Precision Medicine And Personalised Preventive Care

Precision medicine and personalized preventive care are transforming healthcare through the provision of disease pre- diction and prevention tailored to individual health information. Unlike the conventional generalized models, AI-based models examine symptoms, medical history, and risk factors and deliver personal health insights. In skin disease prediction and diabetes prediction, machine learning algorithms like Random Forest and Convolutional Neural Networks are also essential for discovering early indications and proposing protective protocols. Utilizing AI-based risk evaluation enables one to obtain forewarnings and recommendations on life routines in an attempt to make one less likely for the progression of disease. The use of tailored preventive care not only improves the accuracy of diagnosis but also contributes to alleviating the burden on healthcare systems by providing timely intervention and better patient outcomes.

D. AI Cost Analysis in the Promotion of Preventive Health

Another advantage of deploying AI in preventive healthcare is its significant economic benefits, particularly in early intervention cost savings. Pre-diagnosis using AI-driven models reduces the need for complex and expensive treatments by enabling early detection of illnesses [8]. For instance, predictive models that use biomarkers to screen individuals at higher risk of diseases, such as diabetes and cardiovascular conditions, facilitate timely, low-cost interventions compared to expensive clinical treatments and emergency procedures [13]. AI-powered predictive analysis enhances patient care by prioritizing high-risk individuals, thereby optimizing resource allocation and reducing unnecessary tests and procedures.

By focusing on specific diseases, healthcare systems can improve efficiency, minimize resource wastage, and lower overall costs [10]. As AI technologies continue to evolve, their implementation costs are expected to decline, increasing the financial sustainability of healthcare services. When applied in preventive care, AI enables healthcare providers to establish cost-effective business models that promote early intervention, enhance patient outcomes, and reduce overall medical expenditures [11].

E. Ethical and Privacy Concerns in AI-driven Preventive Health

AI in preventive healthcare raises critical ethical and privacy concerns, particularly regarding patient confidentiality and data security. Ensuring the accuracy of AI systems requires large volumes of medical data, which often contain sensitive personal details, increasing the risk of privacy breaches if not managed effectively [11]. Unauthorized access or misuse of such information—whether through cyberattacks or algorithmic misinterpretation, these issues pose serious challenges that must be addressed to ensure the secure deployment of AI in preventive healthcare [8].

Algorithmic bias is another major ethical issue in AI-driven healthcare. If AI models are trained on biased datasets, they may produce predictions that disproportionately favor or dis- advantage certain groups, exacerbating healthcare disparities [12]. Additionally, the lack of transparency in AI decision- making, often referred to as the ”black box” problem, reduces trust among patients and healthcare professionals, as they may struggle to interpret or validate AI-generated recommendations

[16]. Addressing these ethical and privacy concerns requires the establishment of stringent regulatory frameworks that enforce data protection, ensure fairness, and enhance transparency, ultimately fostering public confidence in AI-based preventive care [10]. To address these challenges, regulatory bodies must establish standardized ethical guidelines for AI in healthcare. While GDPR and HIPAA ensure data security, AI-specific regulations are essential for algorithmic account- ability, transparency, and bias mitigation. Rigorous validation, including fairness audits and diverse dataset training, can enhance equitable healthcare outcomes. Collaboration among AI developers, healthcare professionals, and policymakers is vital to creating a balanced framework that fosters ethical AI- driven preventive care [12], [16].

III. METHODOLOGY

A. *Problem Definition*

Diabetes and skin diseases are among the most neglected/poorly diagnosed diseases prevalent in majority population of the globe around. Hence we have built a website using which people showing probable symptoms can at least be made aware about their present health status.

B. *Methodology for Skin Disease Detection*

1) *Classification:* The structure of Convolutional Neural Networks (CNNs) is based on the data provided. In our study,

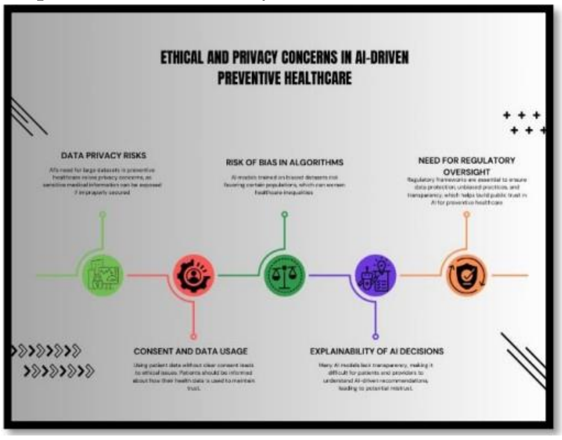

Fig. 2.1. An Image Illustrating Ethical And Privacy Concerns In AI-Driven Preventive Healthcare

we trained and tested a CNN using the Dermnet database, which we modified by incorporating images from personal sources and the internet. The dataset consists of eight classes, with 2500 randomly selected images used for training and 1000 images reserved for pre-training confirmation using ImageNet-pretrained VGG16 models.

After training, we utilized ImageNet-pretrained VGG16 models to evaluate the classification performance. The accuracy of the model was found to exceed 84%.

2) *CNN:* Convolutional Neural Networks (CNNs), a domain-specific subgroup of Artificial Neural Networks (ANNs), are of key importance in deep learning technologies, especially in image and pattern recognition. They are widely used in areas like image and video recognition, facial recognition, medical imaging, natural language processing, and forecasting of time series. In contrast to conventional fully connected Multilayer Perceptrons (MLPs), where every neuron is connected to every neuron in the next layer, CNNs embrace a hierarchical style. This helps to avoid data overload by considering local patterns via convolutional filters. Rather than depending only on weight regularization or dropout strategies for avoiding overfitting, CNNs utilize spatial hierarchies in data and extract complex features through cascaded convolutional layers. Since their effectiveness at extracting and learning efficient hierarchical representations, CNNs are a staple feature in deep architectures nowadays.

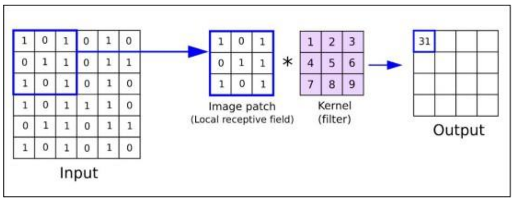

Fig. 3.1. CNN Explanation

Strides determines the sludge movement; if you assign stride
= 1, which is the dereliction, the kernel moves one step at a time. Usually, the sludge size is smaller than the input facts, and the type of addition used between the sludge and the enter records pattern of the sludge size is the fleck product. A fleck product is an element-aware addition between the sludge weights and the sludge size trend of the enter facts, merged into one cost. The sludge length is made lower than the enter information because it allows the same set of sludge weights to be multiplied several times by the enter matrix at particular locations within the photo.

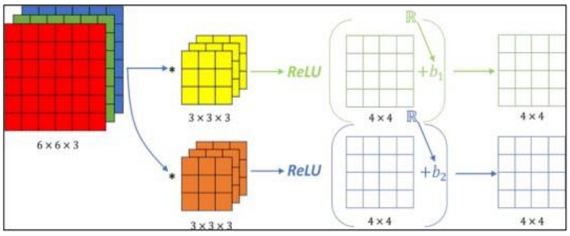

Fig. 3.2. Stride

3) *Maxpooling:* Max pooling is a downsampling method applied in Convolutional Neural Networks (CNNs) to de- crease the dimensionality of feature maps while preserving the most important information. It works by applying a filter to parts of the input data and taking the maximum value from each area, thus preserving key features while reducing computational complexity. This operation refines the pattern recognition ability of the model by extracting the strongest features, filtering out sensitivity towards slight variations in input data. Max pooling, by retaining only the most important spatial information, enhances the efficiency and generalization capacity of CNNs and makes them more robust for image and pattern recognition applications. Maxpooling on a 4 * 4 channel utilizing a 2 * 2 kernel and a step of 2 How we fold using a 2 * 2kernel.The channel has four values 8, 3, If we check the number one 2 * 2 set that the kernel operates on. MaxPooling chooses the greatest figure of this pool, which is"8". Right away, in our space, we're appropriate to generate a kernel which strengthens the cat s eye picture to such a degree that indeed after maximum pooling, the prominent statistics isn't lost. Now while Max Pooling blossoms my pixels, 25 of the remaining pixels is sufficient to maintain the statistics regarding the cat. So there may be a channel or function chart as a means to comprise the cat's eye statistics, anyway of what, to reduce pixels to seventy five. In another instance, we're appropriate to state that we're donating statistics we do not require through building kernels that enable it feasible to acquire the important statistics using Max Pooling.

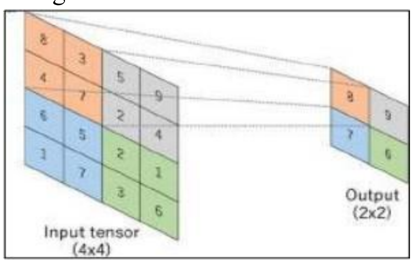

Fig. 3.3. MaxPolling

4) *Residual Block:* ResNet (Residual Network) overcomes the challenges of vanishing and exploding gradients in deep learning by using skip connections. These connections al- low data to bypass certain layers, ensuring smooth gradient propagation during training. Rather than learning the full transformation function $H(x)$, ResNet simplifies the process by focusing on the residual function $F(x)$, defined as:

$$F(x) = H(x) - x$$

This leads to the final output:

$$H(x) = F(x) + x$$

In this approach, $F(x)$ represents the part of the transformation that needs to be learned, while the original input x is directly carried forward. This method improves training efficiency, reduces network degradation, and enables the development of very deep neural networks without performance loss.

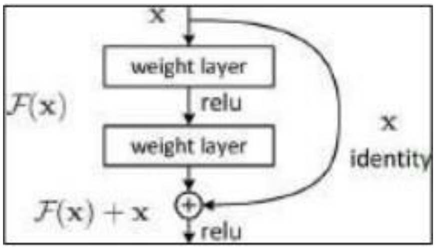

Fig. 3.4. Residual Block

5) *VGG16:* VGG16 (Visual Geometry Group 16) is a widely used deep Convolutional Neural Network (CNN) de- signed for image recognition tasks. The term "deep" refers to the number of layers, with VGG-16 consisting of 16 convolutional layers and VGG-19 containing 19 convolutional layers.

This architecture serves as the foundation for many advanced object recognition models. As a deep neural network, VGGNet has outperformed baseline models across multiple tasks and datasets, including ImageNet. Even today, it remains one of the most powerful and widely used architectures in image processing and computer vision.

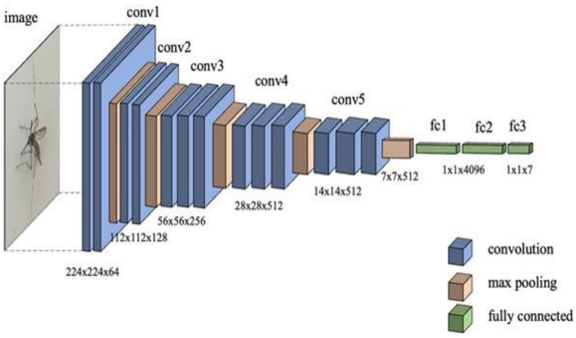

Fig. 3.5. VGG16 Architecture

The VGG16 model performed with 86% accuracy on our dataset. VGG16 is a CNN structure with 13 convolutional and 3 fully connected layers. VGG16 is pretrained using the ImageNet dataset, and we applied transfer learning to leverage it for our particular problem.

For categorization, we incorporated additional custom layers into the pretrained model to make predictions of input into 8 categories. The parameters of the model can also be fine-tuned for improved performance and to obtain the desired output.

C. Methodology for diabetes prediction

The dataset was tried on three different algorithms and the algorithms with most accurate results is selected to predict the final answer. Na¨ıve Bayes was able to achieve 86% of accuracy but Random forest attained maximum accuracy of 94%.

1) Naive Bayes: Naive bayes is wide utilized ai classifier and probabilistic calculation essential uses of naive bayes region unit to channel spam order archives and so on the component feed into the model is independent of each unique that is renascent the value of any of the other component utilized inside the calculation naive bayes enjoys vital benefit is that we've an adapted to confront measure ready to coded until anticipate the yield still speedy it's climbable only and old calculation is equally most reasonable option for planet uses that region unit required to respond to client as by and by as possible Let's Take an Example: You have collection of reviews and classification.

Table 3.1: Naive bayes

Sr. No	Text	Class
1	I loved the movie	+
2	I hated the movie	-
3	A great movie. Good movie	+
4	Poor acting	-
5	Great acting. A good movie	+

Above table define movie review with sentiment data. In on top of table there's a text column that is input and there are 10 unique words that are: - "I, loved, the, movie, hated, a, great, poor, acting, good". categories contain the sentiment information that's negative and positive. which define that movie review is negative or positive based on 10 unique words. Then we've got to convert above data table into features and based on that we tend to get sentiment output. 1st we have to convert it into matrix type and also have to find how many times has that word come back. In table 3.2 unique words are comeback that is "I" word is continual in initial and second review then "Loved" word is repeated in exactly initial review and so on. in class column there are total 5 categories that's mixture of positive and negative classes in this there are 3 positive categories and a couple of negative categories.

Table 3.2: Naive bayes

Sr. No.	I	Loved	the	movie	hated	a	great	poor	acting	good	Class
1	1	1	1	1							+
2	1		1	1	1						-
3				2		1	1			1	+
4				1				1	1		-
5						1	1		1	1	+

Then we've to add up all the positive unique words. Then we've to find the probability against positive category thus the probability is 3/5. Then we computing p(I) prior to that there's a formula that we've to refer that's we add 1 in every probability thus the chance, such as P (*class* | *text*) can never be zero. we are attempting to find out whether a data row must be labeled as negative or positive. because of these we can disregard the divisor. hence we need to compute the probabilities of each classification and hence the probabilities of each feature belonging to each classification. now we need to insert values in our formula In $P\ (I\ |\ +)$ nk 1 i.e., I is happened in initial document only once. This technique is similar for negative text as well. But vocabulary size is same in both the scenarios. Now we have to train our classifier, for that there's one formula that is Vnb = argmax(summation of all the words occur in sentence) positive or negative. thus we take all unique words and then put it in above equation. If $V_j = +$, then:

$$p(+)P\ (I|+)P\ (I|+)P\ (I|+)P\ (I|+)P\ (I|+)\ =\ 6.03\ \times 10^{-7}$$

If $V_j = -$, then:

$$p(-)P\ (I|-)P\ (I|-)P\ (I|-)P\ (I|-)P\ (I|-)\ =\ 1.22\ \times 10^{-5}$$

So get the probability of having this sentence in positive and negative. For positive classification, we get $6.03\ \times 10^{-7}$ and for negative classification, we get 1.22×10^{-5}.

Now, we have to determine whether or not the sentence is classed into positive or negative. So, the answer is negative because of 10^{-5}. In negative values, the min negative range is greater than the most negative number, so that sentence gets classified into the negative class.

D. Random Forest

Each decision tree has high variance, but when multiple trees are combined in parallel, the overall variance is reduced. This is because each tree is trained on a different sample, ensuring that the final prediction is not dependent on a single tree but rather an ensemble of decision trees.

Random Forest is an ensemble learning technique used for both regression and classification tasks. It leverages multiple decision trees and employs a method known as Bootstrap and Aggregation (bagging) to improve accuracy and reduce overfitting.

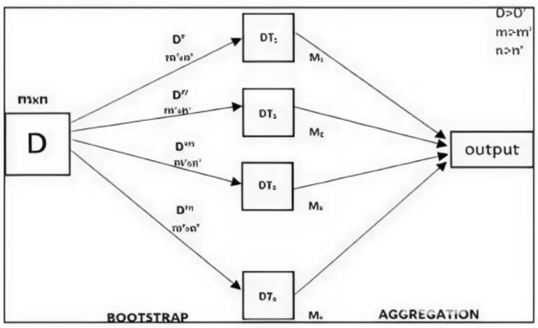

Fig. 3.6. Random Forest

The basic concept of Random Forest is to employ an ensemble of multiple decision trees rather than a single tree for prediction. It is made up of many decision trees that serve as base models. Bootstrap sampling is employed to generate various subsets of the dataset by random row and feature sampling for each tree. Random Forest model had a 94% accuracy. It operates by dividing the dataset into subsets, performing the decision tree algorithm on each subset, and then combining the predictions through majority vote for classification problems.

In our diabetes dataset, every symptom is a binary class (yes or no) for 14 attributes. The information gain is calculated by the decision tree algorithm for every attribute and chooses the one with the maximum gain as the root node. This method is very efficient and simple to implement in Python.

TABLE I ACCURACY

	Na¨ıve Bayes	Random Forest
ACCURACY	86%	94%

IV. RESULT

The system to be proposed is predicting diseases according to the selected disease. When skin disease is chosen model classify the input skin image as ’Not a Skin disease” or any of the seven diseases namely Acne, Hair loss, Melanoma, Poison Ivy and other contact diseases, Nail fungus, Vitiligo, Warts Molluscum and other viral infections using deep learning techniques. The dataset is divided into training and testing sets for model evaluation, and deep learning models are constructed using CNN along with a pre-trained network called VGG16. The accuracy of model is approximately 86%. When diabetes disease is chosen model predicts the class of non-diabetic and diabetic patients using the responses to the questions provided as input by user. This model employs machine learning technique to make the right class prediction. The level of accuracy is over 94%.

V. DISCUSSION

A. Interpretation of Results

The proposed system predicts diseases based on the selected category. The results for each category are as follows:

1) Skin Disease Prediction:

- The model classifies the given skin image as either ”Not a Skin Disease” or one of the seven specific conditions: Acne, Hair Loss, Melanoma, Poison Ivy and Other Contact Diseases, Nail Fungus, Vitiligo, and Warts/Molluscum & Other Viral Infections.
- This categorization is performed through deep learning methods, specifically Convolutional Neural Networks (CNNs) and the pretrained VGG16 model.
- The precision of this model is about 86%, and it is hence effective in detecting different skin diseases.

2) Diabetes Prediction:

- The model predicts whether a patient is non-diabetic or diabetic based on responses to a set of predefined questions.
- This classification is performed using machine learning techniques, which analyze the input data to determine the correct category.
- The accuracy of this model is more than 94%, showcasing its reliability in identifying diabetic patients.

B. Challenges and Limitations

However, the integration of AI in preventive health care runs with trials like data privacy, algorithmic bias, and availability to technology. Data privacy continues to be an issue because these AI models necessitate large volumes of information on patients' health, which attracts insecurity and cases of unlawful entry. There is always a requirement to properly safeguard a patient's identity and the legal requirement to meet laws such as GDPR and HIPAA. Algorithmic bias is another consideration, as creators who employ AI trained on a limited or biased data set often end up programming bias into the results, mostly affecting a specific group of people. This leads to inequity in the access to health care, and hence it cannot be possible to have one type of preventive health for everyone. But, this is an important area because problems regarding the making use of these advanced AI technologies may be a problem in the less developed regions or where there are inadequate resources, hence putting in place a digital divide in relation to access of the healthcare services. These limitations must be solved by collaboration between industries, regulation, and the development of new technologies to help make AI a useful and fair tool to improve preventive healthcare.

C. *Recommendations*

To make AI-based disease forecasting for skin diseases and diabetes more effective, there are a number of important strategies that need to be adopted. Policymakers need to enact clear legal policies to facilitate adherence to data protection laws like GDPR and HIPAA, protecting patients' health information from unauthorized parties. Moreover, addressing algorithmic bias is important since AI models need to be trained on diverse data sets that reflect different ethnicities, skin colors, and age groups to guarantee fairness in diagnosis. Healthcare organizations and developers must also employ bias detection and mitigation strategies to ensure fair healthcare access.

Additional enhancements to model performance and accuracy can be obtained with larger, high-quality datasets fine- tuning deep learning models (CNN, VGG16). Explainable AI (XAI) integration will improve trust by enabling medical practitioners and users to see how predictions are generated. The collaboration of AI developers, dermatologists, endocrinologists, and ethicists is necessary to validate AI models in terms of meeting real-world clinical requirements and their smooth integration into clinical practice.

Extending accessibility and research in ethical AI is also important. AI prediction models need to be integrated into telemedicine and mobile health to reach remote and under- served populations. Greater investment in healthcare research using AI will make models more precise, able to detect more diseases, and yet preserve ethics. By incorporating these suggestions, AI-disease predictive systems can become more accurate, equitable, and accessible to the masses, Ultimately, this contributes to early diagnosis, improved patient outcomes, and more effective preventive healthcare interventions.

VI. CONCLUSION

A. *Key Points*

The project combines CNN for the detection of skin disease and Random Forest for the prediction of diabetes efficiently, providing precise and timely diagnosis. The creation of a web platform greatly enhances the accessibility of healthcare by offering a cost-effective and user-friendly solution for early disease prediction. The system also improves predictive precision with continuous learning and real-time data processing, rendering diagnostic outcomes more accurate. In addition to individual users, healthcare practitioners are able to leverage the system as a decision-making tool, providing useful information about disease patterns and assisting in the proper management of patients. The project is also scalable and extendable, and there is future potential for the integration of the system with healthcare databases, wearable technology, and other disease-detection models, further enhancing the system's application in preventive care.

B. *Future Directions*

The potential advantages of high-level dermatological re- search solutions are vast, with great benefits in minimizing repetitive work for dermatologists and alleviating the workload for medical services. With ongoing development in computer science and medicine, deep learning is developing very fast, making it a promising method for automated skin disease diagnosis. The growing access to inexpensive software packages and high-performance hardware facilitates effective collection and processing of huge datasets, further enhancing the viability of diagnostic systems based on deep learning.

Future studies need to concentrate on creating a generalized skin classification system by improving deep learning models like CNN, VGG16, and Inceptionv3. Although CNN was good at training data but poor at test data, accuracy can be improved with a more diverse and adaptable training dataset with a larger sample size. Also, VGG16 has shown better accuracy than other networks, which suggests its potential for further improvement in skin disease classification.

Another important avenue for future research is the effect of varying skin color on lesion detection and accuracy of classification. Our method is tailored to one skin color, but the effectiveness of the model needs to be verified across a variety of skin types. Increasing the scope of the dataset to include different ethnic groups will enhance the generalizability of AI- based dermatological diagnoses, making it a more universal and credible healthcare tool for a worldwide population.

REFERENCES

[1] G.H. Dayan, et al. Recent resurgence of mumps in the United States. New England Journal of Medicine, vol. 358, no. 15, pages 1580–1589, 2008. https://doi.org/10.1056/NEJMoa0706589. is it correct

[2] S. Jowett and T. Ryan. Skin disease and disability: analysis of the impact of skin conditions, vol. 20, no. 4, pp. 425-429, 1985.

[3] H. Zeng, H. Lui, C. MacAulay, B. Palcic, D. I. McLean. Resources and methods related to skin diagnostic programs, U.S. Patent No. 6,069,689, 2000.

[4] Y. Demchenko, C. De Laat, and P. Membrey. Explaining the structural components of the Big Data Ecosystem, 2014 International Conference on Collaboration Technologies and Systems (CTS), pp. 104-112, 2014.

[5] X. Jin, B. W. Wah, X. Cheng, and Y. Wang. Significance and challenges of big data research, Big Data Research vol. 2, no. 2, pages 59-64, 2015.

[6] P. Jain, N. Pathak, P. Tapashetti, and A. S. Umesh. Confidentiality keeping data decision analysis tree based on single sample selection and decomposition, 2013 9th International Conference on Information Secu- rity and Security (IAS), pp. 91-95, 2013. https://doi.org/10.1109/ISIAS. 2013.6947739 Jessica Velasco et al., International Journal of Advanced Styles in Computer Science and Engineering, 8 (5), September - October 2019, 2632- 2637 2637

[7] J. Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, and G.-Z. Yang. Big health data, IEEE journal biomedical and health informatics, vol. 19, no. 4, pages 1193-1208, 2014.

[8] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. J. Patil, and D. Barton. Big data: management change, Harvard Business Review, vol. 90, when. 10, pages 60-68, 2012.

[9] H. Kantarjian and P. P. Yu. Artificial intelligence, big data, and cancer, JAMA oncology, vol. 1, no. 5, pages 573-574, 2015.

[10] T. B. Murdoch, and A. S. Detsky. The Inevitable Use of Big Data in Health Care, Jama, vol. 309, no. 13, pages 1351-1352, 2013.

[11] A. L. Beam, and I. S. Kohane. Big data and machine learning in health care, Jama, vol. 319, no. 13, pages 1317-1318, 2018

[12] D. D. Miller and E. W. Brown. Artificial Intelligence in the medical profession: the answer to the question ?, The American Journal of Medicine, vol. 131, no. 2, pages 129-133, 2018.

[13] S. Jha no-E. J. Topol. Practicing artificial intelligence: radiologists and pathologists as experts, Jama, vol. 316, no. 22, pages 2353-2354, 2016.

[14] L. K. Tolentino, R. M. Aragon, W. R. Tibayan, A. Alvisor, P. G. Palisoc, and G. Terte. Diagnosis of Surrounding Fingernails Through Artificial Neural Network, Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 10, no. 1-4, pages 181- 188, 2018.

[15] L. C. De Guzman, R. P. C. Maglaque, V. M. B. Torres, S. P. A. Zapido, and M. O. Cordel. Design and Examination of Multi-model, Multi-level Artificial Neural Network for Eczema Skin Lesion Detection, at the Third International Conference on Artificial Intelligence, Modeling and Simulation (AIMS), pages 42-47, 2015.

[16] Singh, A. P., Saxena, R., Saxena, S., & Maurya, N. K. (2024). Artificial intelligence revolution in healthcare: Transforming diagnosis, treatment, and patient care. Asian Journal of Advances in Research, 7(1), 241–263.

[17] Sarker, M. (2024). Revolutionizing healthcare: The role of machine learning in the health sector. Journal of Artificial Intelligence General Science (JAIGS), 2(1), 36–61.

[18] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities, and challenges. Briefings in Bioinformatics, 19(6), 1236–1246.
[19] Shahid, A. H., & Khattak, W. A. (2022). Improving patient care with machine learning: A game-changer for healthcare. Applied Research in Artificial Intelligence and Cloud Computing, 5(1), 150–163.
[20] Parimi, S. S. R. (2020). Research on the application of SAP's AI and machine learning solutions in diagnosing diseases and suggesting treatment protocols. International Journal of Innovations in Engineering Research and Technology, 7(2), 72–81.