Convolutional neural networks over the K-Nearest Neighbor algorithm are used in a novel healthcare system.

Ashutosh Tripathi[1], T. P. Anithaashri [2*]

[1]Research Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. Saveetha University, Chennai, Tamilnadu, India, Pincode:602105

[2*]Project Guide, Corresponding Author, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Saveetha University, Chennai, Tamil Nadu, India, Pincode 602105.

Abstract

Aim: To develop a novel healthcare system for the detection of diseases in humans using blood cells by Convolutional Neural Networks (CNN) algorithm over K-Nearest Neighbor algorithm. **Materials & Methods:** Image processing can be defined as the technical analysis of an image by using complex algorithms. by using CNN method the image recognition was done and performed with the sample size of 80 each and the software tool jupyter notebook is used. Accuracy values for identification of disease detection and face masks are calculated to quantify the performance of the CNN algorithm against the k-nearest neighbor algorithm method. Accuracy values for identification of disease detection are calculated to quantify the performance. **Results:** The analysis on the training dataset and test dataset has been performed successfully using SPSS and acquired accuracy (92.3%) for the CNN algorithm compared to KNN algorithm, which gave (88.8%) accuracy with a statistically insignificant difference between the two groups ($p=0.003$; $p<0.05$) with confidence interval 95%. The resultant data depicts the reliability in independent sample tests. **Conclusion:** On the whole process of prediction of accuracy, the CNN model gives significantly better performance compared with the k-nearest neighbor model.

Keywords: Disease Detection, Convolutional Neural Network, Deep Learning, Face mask Detection, K-Nearest Neighbor Algorithm, Novel Healthcare System.

Introduction

This project was created to provide safe, reliable, time saving efficient, comfortable and affordable application services for people after the outbreak of the worldwide pandemic COVID-19 to detect face masks and diseases like malaria, pneumonia, breast cancer, skin cancer using deep learning methods is seen as having accomplished this objective due to the drawbacks in the existing systems. The idea was motivated by authors (Yadav and Jadhav 2019)and (Meivel et al. 2021) to develop a disease detection system with the images uploaded by the users and person wearing face mask or not in a live stream video. A smart application was developed to combine the both ideas by using CNN multi-view classification (LeCun, Bengio, and Hinton 2015) in the area of machine learning as one way to improve classification performance. This approach was taken to anticipate the need for creating and maintaining the application on multiple operating system platforms and was expected to cut development time frame while retaining consistent interface and still enable platform specific's feature usage.

In the existing models the available technological options for developing multi-platform applications are web-based apps and hybrid apps. Hybrid apps, though also relying on standardized web technologies, are bundled within a native app "container", which serves as a bridge to access device hardware and functions like, using the camera, sensor, memory access and interface. (Peppers, Taskos, and Bilgin 2016) There are several hybrid multi-platform software applications that have been developed. These frameworks let a developer use one programming language to build an multi application that supports one with platform independency (Awan et al. 2014). Existing multi-platform framework could not decrease the effort of devising appropriate user interfaces for the different device types and orientations, however it still helped to create a codebase which is usable across multiple platforms. (Freire, de Abreu Freire, and Painho 2014). In machine-learning-based image classification infers an object's class from just a single image showing an object, especially for challenging classification problems, the visual information conveyed by a single image may be insufficient for an accurate decision. The work proposed is an approach to develop a smart application to provide services of health disorder detection and covid-19 face mask detection system to the users into a single application. The application is developed in an hybrid framework using (PLOS ONE Staff 2021; Seeland and Mäder 2021) multi-view Convolutional Neural Networks to extract and encode visual features from the dataset images trained. (Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020)

Previously our team has a rich experience in working on various research projects across multiple disciplines(Venu and Appavu 2021; Gudipaneni et al. 2020; Sivasamy, Venugopal, and Espinoza-González 2020; Sathish et al. 2020; Reddy et al. 2020; Sathish and Karthick 2020; Benin et al. 2020; Nalini, Selvaraj, and Kumar 2020).The research gap in the existing system (Chavda et al. 2021) pays way to development of a smart application for health disease detection system along with face mask detection system and provides the generic solution (Cheon 2019; Yamashita et al. 2018; Tripathi and Kumar 2019)to the problem of cross platform multipurpose application development. The objective of the proposed system is to overcome the difficulties of these research gaps using Convolutional Neural Network algorithm and k-nearest neighbor algorithm. The aim is to implement a system by using Convolutional Neural Network algorithms which can detect the presence of the diseases & masks efficiently and provide a prediction of it with high accuracy in image processing by data augmentation in image classification

Materials And Methods

The Study setting of the proposed work is done in the Compiler Design Lab, Saveetha School of Engineering, Saveetha University. The number of groups identified are two. Group 1 is the Convolutional Neural Network and group 2 is the K-Nearest Neighbor Algorithm. Convolutional Neural Network method was iterated various number of times with a sample size of 80 calculated from clinical website and SPSS analysis is carried out with level of significance ($p<0.05$).

Convolutional Neural Network Algorithm

The architecture of a CNN is designed to take advantage of the 2D structure of an input in the image processing. This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. CNN's are easier to train and have fewer parameters than fully connected networks with the same number of hidden units. CNN does the image classification based on features such as pixel intensity and image edges. The following are the steps of the Convolutional Neural Network algorithm.

Pseudo Code for Convolutional Neural Network Algorithm

Step 1: Start the process
Step 2: Load datasets path
Step 3: Prepare dataset for training
Step 4: shuffle the data
Step 5: assigning labels and features
Step 6: Normalizing X and converting labels to categorical data
Step 7: Split X and Y for use in CNN
Step 7: Define, compile and train the CNN Model
Step 8: Accuracy and Score of model
Step 9: Stop the process

Novel K-Nearest NeighborAlgorithm

It is a supervised machine learning algorithm. The algorithm can be used to solve both classification and regression problem statements. The K-NN algorithm assumes the similarity between the new case/data and available cases and puts the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suited category by using K-NN algorithm. K-NN algorithm uses 'feature similarity' to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. The following are the steps of k-nearest neighbor algorithm :

Pseudocode for Novel K-Nearest NeighborAlgorithm

Step 1: Start the process.
Step 2: Import libraries and dataset
Step 3: Split dataset into test and train split.

Step 4: Train the model and run predictions on the test split data.
Step 5: Evaluate algorithm and check the validation error, run again with different K values and consider the k value that has the minimum validation error.
Step 8: Accuracy and Score of model
Step 9: Stop the process

Anaconda navigator is used for execution of the project code. It helps to manage and access notebook files and any kind of python files. By giving the python environment a command prompt can provide easy access to the code and execution. Main tools that need to be installed in the python environment are keras and tensorflow. Minimum of 4GB RAM is required to compile and execute the project code. Preferred operating systems are windows and ubuntu. The above mentioned method is for users using windows OS. Using anaconda navigator software and anaconda prompt that install the necessary modules. The GUI for the application is created using python Tkinter.

Statistical Analysis

To check the data accuracy & reliability an SPSS a statistical software is used with a default alpha value of 0.05 for the sample size of 80. The independent variables for the dataset were blur, varying lighting condition,shadowing effects,image size of the images. Many potential variables are dependent in image classification like spectral signatures, vegetation indices, transformed images, textural or contextual information, multitemporal images, multisensor images, and ancillary data. The image is segmented and binarized to build the function that contains the interest area for detection. The bar graph and the error graph were generated for comparison of differences between the CNN and K-NN algorithms.

Results

Table1 inferred Correlations for Study. Data collection from N=80 sample datasets for Convolutional Neural Network (92.3%) Compared with K-Nearest neighbor (88.8%) using target variable as Independent variable.

Table 2 represents the sample size (N=80) (the table consists of sample values between the range of 90-100 for accurate values therefore N=10), Mean, Standard deviation and Standard error mean are classified based on the accuracy and loss of the data. The accuracy 95% of CNN is significantly higher compared to the K-Nearest Neighbor algorithm.

Table 3 represents the significance of the data and standard error difference, where significance of Convolutional Neural Network (CNN) and K-Nearest neighbor with the confidence interval as 95% and level of significance of 0. 05.

Figure 1 represents the analysis of the accuracy of Convolutional Neural Networks (CNN) and K-Nearest neighbor for better improvisation in the application . Figure 2 represents the comparison of mean accuracy of Convolutional Neural Networks (CNN) The comparison of accuracy gained. The accuracy of group 1 is 95% and group 2 is 93%. The has significantly performed better when compared to K-Nearest neighbor Group 1 appears to produce the most consistent results with its standard deviation ranging from the lower 93's to higher 95's. Group 2 appears to produce the most variable results with its standard deviation ranging from 85's to 90's. There is a significant difference between (CNN) and (K-NN)

Discussion

In this study of a smart application for health disorder discrimination system and face mask detection the Convolutional Neural Networks have higher accuracy of 95% in comparison to K-Nearest Neighbour algorithm having accuracy of 93%. Convolutional Neural Networks have better significance ($p<0.05$) than K-nearest neighbor and while using the independent sample t-tests.

The similar work is carried out by the author (Adinugroho, Heriyanni, and Musthofa 2017) the mobile application development was done by combining native mobile technology and web technology using WebView API. This approach was taken to anticipate the need for creating and maintaining the application on multiple mobile platforms and was expected to cut development time frame while retaining consistent interface and still enable platform specific's feature usages. The further development was done by the (*Proceedings of the 19th International*

Conference on World Wide Web - WWW '10 2010)assigning an SLA only to the front-end service. Other services were not given any particular response time objectives. Services were autonomously responsible for their own provisioning operations and collaboratively negotiated performance objectives with each other to decide the provisioning services. The barrier in developing such an application was that an entire multi-service application to perform better and good under varying workloads to meet the functionalities and improve resource utilization. The limitations faced were that there are many approaches in developing a cross platform native application but issues are with the tools used is that compilers need to transform the source code written in high-level programming language to an executable program for any system to run and support the application at an user interface level. (El-Kassas et al. 2017)

Inorder to overcome the above drawbacks in the proposed system and in the application CNN algorithm is used. CNN-based methods have various strategies to increase the performance of image classification on datasets. A hybrid approach offers a simple solution for developing applications for devices by writing the code once and deploying it to different operational systems will help users to quickly access the health disorders and face mask detection application and reduce maintenance and save time. Hybrid structures are suitable options for the real time benefits in the use of applications for business or for education. (Andrade et al. 2015). The proposed work uses . Convolutional neural networks as it can create "lightweight" and compact frameworks for development of applications. They allow reducing the size of the model significantly as well as the time of cycle processing. Nevertheless, they guarantee the accuracy of recognition and mitigate the negative effects of unbalanced data on recognition results. The reusability of source codes , maintenance of screen sizes are left for future works

The results of the proposed algorithm are better in both experimental and statistical analysis, there are certain limitations in the work. The main limitation in this experiment is that the attributes in the dataset contain very few to predict accuracy (%) of the disease detection and faces as the image considered are clear pixels. Further research should consider other deep learning algorithms to ascertain more efficient ways to perform better for large datasets. It is recommended that further research should be carried out on other parameters that can improve the accuracy of detection.

Conclusion

It is inferred that the accuracy in detection of disease through extracted images of blood cells and face mask detection using Convolutional Neural Network algorithm gives significantly better results than K-Nearest neighbor algorithm. CNN shows the accuracy of 92.3 % using and KNN shows the accuracy of 88.8 % in detection.

Declarations

Conflict of interests

No conflict of interest in this manuscript.

Author Contribution

Author AT was involved in data collection, data analysis, and manuscript writing. Author TPA was involved in conceptualization, data validation and critical review of manuscript.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete this study:

1. MANAC Infotech pvt.Ltd., Hyderabad.
2. Saveetha School of Engineering.
3. Saveetha University.
4. Saveetha Institute of Medical Technical Sciences. .

References

1. Adinugroho, Timothy Yudi, Eileen Heriyanni, and Tegar Aryo Sulthon Musthofa. 2017. "Mobile Web Application for Students' and Lecturers' Portfolio Gallery." *Library Hi Tech News*. https://doi.org/10.1108/lhtn-02-2017-0012.
2. Andrade, Paulo R. M. de, Paulo R. M. de Andrade, Adriano B. Albuquerque, Otávio F. Frota, Robson V. Silveira, and Fátima A. da Silva. 2015. "Cross Platform App : A Comparative Study." *International Journal of Computer Science and Information Technology*. https://doi.org/10.5121/ijcsit.2015.7104.
3. Awan, Irfan, Muhammad Younas, Xavier Franch, and Carme Quer. 2014. *Mobile Web Information Systems: 11th International Conference, MobiWIS 2014, Barcelona, Spain, August 27-29, 2014. Proceedings*. Springer.
4. Benin, S. R., S. Kannan, Renjin J. Bright, and A. Jacob Moses. 2020. "A Review on Mechanical Characterization of Polymer Matrix Composites & Its Effects Reinforced with Various Natural Fibres." *Materials Today: Proceedings* 33 (January): 798–805.
5. Chavda, Amit, Jason Dsouza, Sumeet Badgujar, and Ankit Damani. 2021. "Multi-Stage CNN Architecture for Face Mask Detection." *2021 6th International Conference for Convergence in Technology (I2CT)*. https://doi.org/10.1109/i2ct51068.2021.9418207.
6. Cheon, Yoonsik. 2019. "Multiplatform Application Development for Android and Java." *2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)*. https://doi.org/10.1109/sera.2019.8886800.
7. El-Kassas, Wafaa S., Bassem A. Abdullah, Ahmed H. Yousef, and Ayman M. Wahba. 2017. "Taxonomy of Cross-Platform Mobile Applications Development Approaches." *Ain Shams Engineering Journal*. https://doi.org/10.1016/j.asej.2015.08.004.
8. Freire, C. E. de Abreu, C. E. de Abreu Freire, and Marco Painho. 2014. "Development of a Mobile Mapping Solution for Spatial Data Collection Using Open-Source Technologies." *Procedia Technology*. https://doi.org/10.1016/j.protcy.2014.10.115.
9. Gudipaneni, Ravi Kumar, Mohammad Khursheed Alam, Santosh R. Patil, and Mohmed Isaqali Karobari. 2020. "Measurement of the Maximum Occlusal Bite Force and Its Relation to the Caries Spectrum of First Permanent Molars in Early Permanent Dentition." *The Journal of Clinical Pediatric Dentistry* 44 (6): 423–28.
10. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44.
11. Meivel, S., K. Indira Devi, S. Uma Maheswari, and J. Vijaya Menaka. 2021. "Real Time Data Analysis of

Face Mask Detection and Social Distance Measurement Using Matlab." *Materials Today. Proceedings*, February. https://doi.org/10.1016/j.matpr.2020.12.1042.

12. Nalini, Devarajan, Jayaraman Selvaraj, and Ganesan Senthil Kumar. 2020. "Herbal Nutraceuticals: Safe and Potent Therapeutics to Battle Tumor Hypoxia." *Journal of Cancer Research and Clinical Oncology* 146 (1): 1–18.
13. Peppers, Jonathan, George Taskos, and Can Bilgin. 2016. *Xamarin: Cross-Platform Mobile Application Development*. Packt Publishing Ltd.
14. PLOS ONE Staff. 2021. "Correction: Multi-View Classification with Convolutional Neural Networks." *PloS One* 16 (4): e0250190.
15. *Proceedings of the 19th International Conference on World Wide Web - WWW '10*. 2010. New York, New York, USA: ACM Press.
16. Reddy, Poornima, Jogikalmat Krithikadatta, Valarmathi Srinivasan, Sandhya Raghu, and Natanasabapathy Velumurugan. 2020. "Dental Caries Profile and Associated Risk Factors Among Adolescent School Children in an Urban South-Indian City." *Oral Health & Preventive Dentistry* 18 (1): 379–86.
17. Sathish, T., and S. Karthick. 2020. "Gravity Die Casting Based Analysis of Aluminum Alloy with AC4B Nano-Composite." *Materials Today: Proceedings* 33 (January): 2555–58.
18. Sathish, T., D. Bala Subramanian, R. Saravanan, and V. Dhinakaran. 2020. "Experimental Investigation of Temperature Variation on Flat Plate Collector by Using Silicon Carbide as a Nanofluid." In *PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS ENGINEERING: ICRTMME 2019*. AIP Publishing. https://doi.org/10.1063/5.0024965.
19. Seeland, Marco, and Patrick Mäder. 2021. "Multi-View Classification with Convolutional Neural Networks." *PloS One* 16 (1): e0245230.
20. Sivasamy, Ramesh, Potu Venugopal, and Rodrigo Espinoza-González. 2020. "Structure, Electronic Structure, Optical and Magnetic Studies of Double Perovskite Gd2MnFeO6 Nanoparticles: First Principle and Experimental Studies." *Materials Today Communications* 25 (December): 101603.
21. Tripathi, Shyava, and Rishi Kumar. 2019. "Image Classification Using Small Convolutional Neural Network." *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. https://doi.org/10.1109/confluence.2019.8776982.
22. Venu, Harish, and Prabhu Appavu. 2021. "Experimental Studies on the Influence of Zirconium Nanoparticle on Biodiesel–diesel Fuel Blend in CI Engine." *International Journal of Ambient Energy* 42 (14): 1588–94.
23. Yadav, Samir S., and Shivajirao M. Jadhav. 2019. "Deep Convolutional Neural Network Based Medical Image Classification for Disease Diagnosis." *Journal of Big Data*. https://doi.org/10.1186/s40537-019-0276-2.
24. Yamashita, Rikiya, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. 2018. "Convolutional Neural Networks: An Overview and Application in Radiology." *Insights into Imaging* 9 (4): 611–29.

Tables And Figures

Table 1. Correlations for Study. Data collection from N=10 sample datasets for Convolutional Neural Network (92.3%) Compared with K-Nearest neighbor (88.8%) using target variable as Independent variable

Sample No.	Convolutional Neural Network	K-Nearest neighbor
1	92.3	88.8
2	92.6	90.4
3	93.0	90.8
4	93..4	91.1
5	93. 7	91.6
6	94. 1	91.8
7	94. 4	92.3
8	94. 6	92.7
9	94.8	92.9
10	95. 0	93.0

Table 2. Mean, Standard Deviation, and Standard Error of Precision and Accuracy of CNN and K-NN Algorithms Statistical Analysis The precision and accuracy levels of the two techniques differ statistically significantly. The precision of K-NN is lower (88.81%) than that of CNN, which has a better accuracy (92.36%).

Group	N	Mean	Std. Deviation	Std. Error Mean
Algorithms				
Convolutional Neural Network (CNN)	80	92. 3600	. 68508	. 21664
K-Nearest neighbor.	80	88. 8060	. 69570	. 22000

Table 3 : Comparison between Convolutional Neural Network and K-Nearest Neighbour with N=10 samples of the dataset with the highest performance of 92.36 and 88.81 in the sample (when N=1) using data size=80 and the 70% training data and 30 % testing data

		F	**sig.**	**t**	**df**	**sig. (2-tailed)**	**Mean difference**	**Std. Error Difference**	**95% Confidence interval of the difference Lower**	**95% Confidence interval of the difference Upper**
Accuracy	Equal variance assumed	3.27	0. 0292	15. 8	18	. 003	4. 90	. 3087	4. 29132	5. 58868

Accuracy	Equal variance not assumed s			15. 8	17. 9	. 003	4. 94	. 3087	4. 29131	5. 58869

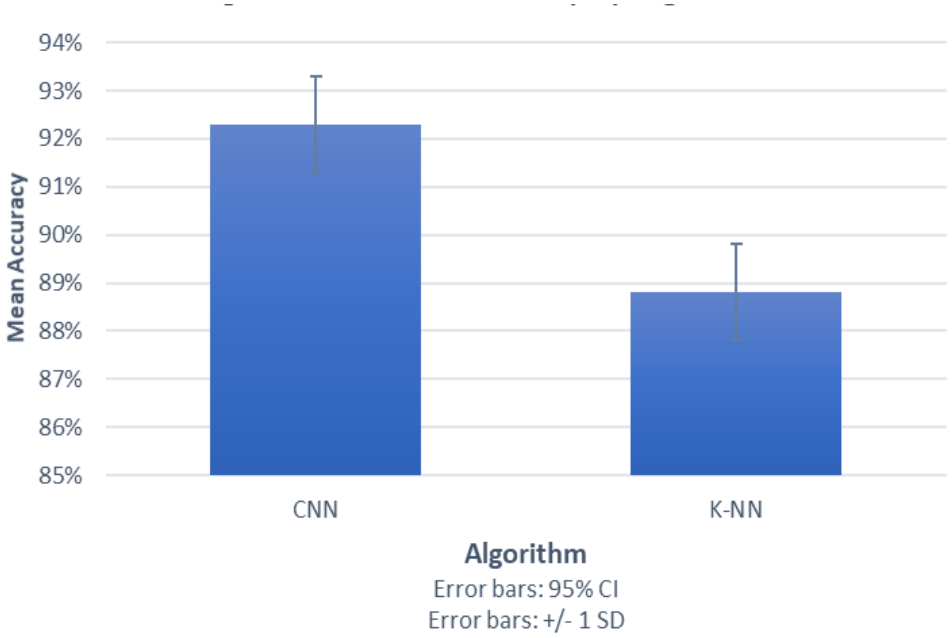

Fig. 1. Comparisons of Convolutional Neural Networks (CNN) that appear to produce the most consistent results with minimal standard deviations. There is a significant difference between Convolutional Neural Network (CNN) and Support vector machine. X Axis: Convolutional Neural Network (CNN) vs Support vector machine (SVM) Y Axis: Mean accuracy of detection ± 1 SD.