ENSEMBLE OF DEEP LEARNING MODELS FOR IRIS RECOGNITION

Mohammed Hafeez M.K[1*], M Sharmila Kumari[2]

Article History: **Received: July 2022** **Revised: December 2022** **Accepted: February 2023**

Abstract

Biometrics representing biological and physiological characteristics of an individual which are used to authenticate individuals are becoming increasingly common in many areas such as banking, airport transfer, document authorization, cyber forensics, land registration etc. Among many biometric traits, iris biometric is considered the most reliable and secure metric due to its universality, uniqueness, permanence, and collectability. It is noted by the researchers that the iris texture also varies between the left and right eye and even between twins, making it a highly secure authentication method. The existing iris recognition methods address many challenges such as occlusion, scaling, rotation, motion blur, pupillary dilation, irregular reflections in non-cooperative environments, etc. In order to improve the accuracy or robustness against these challenges, deep-learning-based person authentication methods have been developed that accurately identify genuine and imposter by analyzing the differences between corresponding patches in pairs of iris images. The convolution neural network (CNN) based VGG 16, Inception, Nasnet, and Mobilenet deep learning models provide the best recognition performance on different iris datasets. Motivated by the fact that the ensemble algorithms significantly improve the accuracy of any computer vision systems, in our work, we have designed an ensemble model that combines the best features of deep learning models. The ensemble model uses techniques such as stack ensemble, voting, and bagging with classifiers such as Logistic regression, random forest, decision tree, and K nearest neighbour to improve the classification performance of the iris recognition system. Our experiments using four well-known datasets (CASIA, Polaris IITD, Ubiris, and MMU) have demonstrated the accuracy of the proposed ensemble model designed with CNN-based deep learning algorithms.

Keywords — Machine Learning, Convolutional Neural Network (CNN), Deep Learning, Ensemble technique, Iris recognition.

[1,2] Department of Computer Science and Engineering, P A College of Engineering, hafiz123ster@gmail.com, sharmilabp@gmail.com

***Corresponding Author:** Mohammed Hafeez M.K
Department of Computer Science and Engineering, P A College of Engineering, hafiz123ster@gmail.com

DOI: 10.48047/ecb/2023.12.si10.00502

I. INTRODUCTION

Biometric systems are currently the most reliable way to ensure secure and authentic transactions for individuals. The iris recognition system stands out among the various biometrics due to its high level of security and reliability. The iris is a visible internal organ with a unique physical structure that is highly protected and contains a large amount of data. It is also highly variable, making it difficult to replicate or forge [1, 2]. However, most iris recognition systems are designed for controlled environments and fail in non-cooperative environments, such as those with noise, off-axis rotations, motion blur, or partially open eyes. Deep-learning frameworks based on feature-learning algorithms and Convolutional Neural Networks are proposed to address these challenges, By extracting effective features from the ImageNet dataset, we can improve the accuracy and performance of iris recognition systems, even in complex environments. We also use transfer learning techniques to reduce training time and improve performance with less data. Additionally, we use an ensemble technique that combines the features of highly performing deep learning models to further improve the overall performance of the iris recognition system. Our experimental study is tested on four public datasets namely Polaris IITD iris database, CASIA-Iris, Ubiris, and MMU. From the experimental results, we can conclude that our approach shows promising results for improving the accuracy and reliability of iris recognition systems in a variety of environments.

II. LITERATURE SURVEY

In (Wang, 2020) [4], the main focus of the research work is to propose a new method for iris recognition that uses dilated residual features to improve accuracy using low-quality images and non-ideal conditions. The use of dilated convolutional kernels and residual network learning helps to capture more contextual information from the iris image and optimize the training process, respectively. This approach is different from traditional methods that rely on handcrafted features or shallow learning algorithms. Overall, the proposed method represents a new approach to iris recognition that can potentially improve its accuracy and generalization capability compared to existing methods.

In (Thakkar, 2020) [3], the Gabor channels used the iris highlights to create the highlight vector. It was focused on iris identification with the introduction of the neural network using vectors. The Hough change was extracted from the edge image using dim-scale iris image. The iris ring can be recovered from the image and transformed into a square image with a resolution of 64x512 pixels by deleting the extraneous elements. A low-pass Gaussian channel is utilized to eliminate disturbance and create a normalized image with a high concentration of recurrence. It is then converted into a 1-D cluster with a vector of 160x1 elements using AAD esteems. This one-dimensional cluster is utilized as a contribution to the complex neural organization for verifying a certified client.

In (Sardar, 2020) [5], a UNet variation based on squeeze modules was proposed to speed-up the iris segmentation process. By reducing the number of boundaries incorporated, this strategy helped to increase capacity efficiency. This classification-based approach made use of the potential and developments of deep learning methods. Over-fitting was avoided because of the existence of predictable model. The time complexity in producing an accurate segmentation is significantly improved.

In (Kuo, 2019) [4], the authors have provided a different way for more accurate iris recognition. The mandatory scale and expansion adjustments during iris imaging were identified, and the crucial hotspot for the frequently observed iris miss occurrences was developed. Convolutional bit fusion was used in the process to learn and obtain high iris coordination. The effectiveness of this methodology has been confirmed by the trial results using three different public iris photo data sets and within the database and cross-information base execution assessment. Iris images naturally reflect visual information and also enhance iris image coordination with accuracy. The Mask Net, which was independently constructed, was used in this methodology.

In (Kerrigan, 2019) [6], as opposed to being open-sourced earlier in an iris recognition setting, a triple open-sourced iris segmentation apparatus that relies on typical inverted neural networks was developed for iris segmentation task. Deep residual networks (He, 2016), CNN combining expanded convolutions and SegNet, where segmentation results of two model iris pictures demonstrated the method of applying the subsequent sporadic segmentation covers to a conventional, Gabor wavelet based iris coordinating, along with the evaluation of the subsequent iris recognition exactness acquired.

Hence, the results of the segmentation of the iris data, which were acquired using a variety of groups and sensor data, including those collected from CASIA Iris and UBIRIS posts Mortem Irises v2 posthumous iris images, clearly demonstrated the performance of the CNN.

In (Khalifa, 2019) [7], a deep convolutional neural network-based iris segmentation technique was developed. Three convolutional layers are used for extraction out of a total of 16 layers, with different convolution window sizes. The suggested approach achieved 98.88% with a significant improvement in the testing stage.

In (Li, 2019) [8], a hybrid approach utilizing edge and self-learning was suggested for iris recognition. A six-layer, well-designed Faster R-CNN was used to find and group the eyes. To organize the pupillary region, a Gaussian combination model with the bouncing box discovered by the Faster R-CNN was proposed. The limbus limit was constructed using five crucial limit points.

Iris recognition and filtering were investigated using a residual convolution neural net that was concurrently trained from the feature representations and performance recognitions. The transfer learning technique was used to create a framework for iris recognition (Minaee, 2019) [9]. Convolutional neural network parameters that had been previously trained were defined using the well-known IIT Delhi iris recognition dataset. The iris database contained 2240 images of the iris that were taken of 224 different people.

In (Bazrafkan, 2018) [10], the main objective was to apply a deep learning strategy to design better segmentation on low-quality iris images. The major goal is to build tests that nearly mirror the real images produced by handheld cameras using the data. By computing the cumulative total with adequate data given to it, the network will efficiently process the raw input. It begins with 3x3 window and an isolated pooling layer that introduces extra noise. Using skip associations, the architecture proves the segmentation of unconstrained, low-quality images captured outside of any restrictions.

In (Zanlorensi, 2018) [11], deep learning models based on VGG and ResNet-50 for iris recognition have been proposed. The authors used interchange gaining from the face space and specialized information increase approach for iris pictures. To boost speculation and avoid over fitting, two Convolutional Neural Network (CNN) models developed for face recognition were kept and utilized for iris identification (or highlights). Only non-standardized and non-fragmented iris images were shown using the methodologies as a contribution from the authors.

In (Karakaya, 2018) [12], a Convolution neural network was used to enhance the performance of off-target iris identification frameworks [20]. Since iris recognition frameworks are less forceful than conventional frameworks, they improve the display of off-angle iris recognition in both conventional and non-conventional iris recognition systems. The main goal of using convolutional neural organizations (CNNs) was to examine the traditional iris identification structure using segmentation, standardization, and CNN-based encoding and coordinating. The researchers employed segmentation-free iris images in unconventional topologies to look at how periocular regions affected vision.

In (Arsalan, 2017) [13], the CNN approach is based on adjusted roundabout HT, which distinguishes the ROI by the slightly increased iris range. The information retrieved from the return on investment was calibrated using VGG. The CNN layer provides two highlights of the yield. Non-iris focuses described in this way, take features into account, to find the actual iris limit. The authors thought about using semantic segmentation organization (SSN), which can use the full image as information, to address these problems. Even in critical circumstances, this technique's poor performance provided an accurate ID of the real limit. This technique's primary phase includes base cap separation, noise expulsion, canny edge locator, contrast upgrade, and modified HT to solve the iris limit. The next stage involved using CNN to suit the actual iris limit, with a picture contribution of 21X21 pixels. A reduction in handling time and iris segmentation error was seen when the second stage segmentation technique was used just inside the ROI defined by the hypothesized iris limit. The SR regions inside the image contribute to CNN and are standardized by the average RGB worth of the iris to reduce the impact of regions in iris segmentation process.

III. PROPOSED METHODOLOGY

Deep Learning is a branch of machine learning that focuses on artificial neural network techniques for simulating the human brain. Deep learning perform well in case there exists a large amount of data to train the system under study. Convolutional neural networks (CNN), a supervised deep learning method, have three component layers: pooling, convolution, and fully connected layer. The convolution layer picks up the features of the image by moving the filter around it in steps of a certain size called stride. It computes the inner product between the pixel values and the parameters by advancing horizontally and vertically on an image depending on the selected stride window until it covers the full image. By learning invariant features and serving as a regularizer, the pooling layer lowers the computational cost, network training time, parameter count, and overfitting. A feed-forward neural network's fully connected layer flattens out the final pooling or convolution layer and completely connects the neurons to the output layer. The performance and training speed of CNNs is directly impacted by the size, number, length, and pooling size of the kernels.

The iris identification framework utilized in this study is based on transfer learning and is optimized using pre-trained convolutional neural networks (trained on ImageNet) on a variety of well-known iris recognition datasets.

MobileNet V2 is a powerful and efficient architecture for mobile computer vision applications, enabling high-quality deep learning models to be deployed on resource-constrained devices. MobileNet achieves its efficiency by using a technique called depthwise separable convolutions, which break down a traditional convolution operation into two separate operations: a depthwise convolution and a pointwise convolution. The depthwise convolution applies a single filter to each input channel, while the pointwise convolution applies a linear combination of 1x1 convolutions to the output of the depthwise convolution. By using depthwise separable convolutions, MobileNet significantly reduces the number of parameters and computations required while maintaining high accuracy on tasks such as image classification and object detection.

The Inception architecture is a powerful and efficient deep learning architecture that has been widely adopted in the computer vision community for image classification, object detection and semantic segmentation. The architectural version of Inception-v3 includes batch normalization, which speeds up training and prevents overfitting, and residual connections enhance gradient flow through the network. The primary innovation of the Inception architecture is the use of several filters of varied sizes within each network layer. This enables the network to simultaneously learn a variety of feature sizes and aspect ratios rather than being limited to a single filter size.

NASNet-Large is a computationally efficient and high-accuracy image classification convolutional neural network architecture consisting of a series of convolutional layers with skip connections and a combination of traditional and depthwise separable convolutions. The NAS cell consists of a combination of convolutional layers, pooling layers, and skip connections, and is designed to be flexible and adaptable to different input sizes and task requirements. The architecture automatically generated and evaluated millions of different architectures with the highest validation accuracy by minimizing the computational cost.

The VGG16 is a simple and elegant architecture consisting of 16 layers of convolutional and fully connected layers. The key feature of VGG16 architecture is that it uses small 3x3 filters for all convolutional layers, with a stride of 1 and padding to maintain the spatial dimensions of the feature maps. To avoid overfitting and to reduce the spatial resolution of the feature maps, the network also includes max-pooling layers with a stride of 2.

The EfficientNet architecture allows the network to learn more complex patterns in the data. It reduces over fitting and improves generalization performance by scaling the depth, width, and resolution of the network simultaneously. EfficientNet has several variations, including EfficientNet-B0 to EfficientNet-B7, with each model having a higher number of blocks with different parameters and computational requirements.

The performance of the deep learning models is evaluated based on training accuracy, validation accuracy, testing accuracy, precision, recall and F1_Score. The performance obtained using the deep learning technique is then compared with the performance of traditional texture-based algorithms like principal component analysis, local binary pattern, the variants of local binary pattern such as robust local binary pattern and local ternary pattern. It is known that the VGG16, Inception, NASNet and MobileNet provide the best performance on the datasets selected. The performance of the model is further improved by ensembling the features of the best-performing CNN models. The ensembling techniques such as stack ensembling, voting and bagging are used to ensemble the model features.

ENSEMBLING LEARNING ARCHITEC TURE

In numerous applications, convolutional neural networks have demonstrated their superiority as a deep learning model. The CNN models have consistently demonstrated their proficiency while handling massive data sets to extract features and make predictions [14]. In the vast majority of applications, a single CNN model is used. Ensembling is the process of merging various learning algorithms to gain their collective performance or to enhance the performance of current models by mixing various models to produce one trustworthy model. The basic motivation for employing ensemble approaches is to build a model with less volatility and bias, which will ultimately produce better predictions. Customized ensemble learning models have noticed an improvement in their effectiveness. The architecture of the ensemble learning model is as shown in fig 1. The ensemble model is created by selecting the best features of the various individual models such as VGG16, Inception, Nasnet and Mobilenet using recursive feature elimination technique. The combined model contains the best features of each model.

The ensembling techniques such as voting, stack ensemble, dirichlet ensemble and bagging are applied to obtain the ensemble model. Each technique implements a set of classification algorithms such as KNN, Decision tree, Logistic regression and Random forest to obtain the best ensemble model. The ensemble model is evaluated based on the accuracy, precision, recall and sensitivity and it is compared with the individual model. The comparative analysis between the individual and ensemble model is presented in the experimental section.

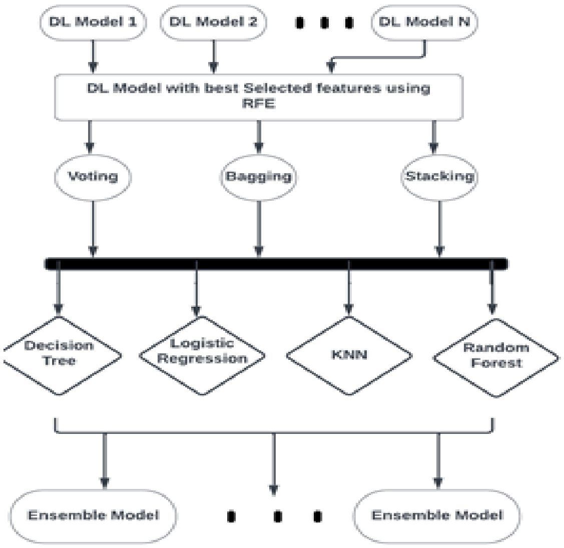


Fig 1: Architecture of Ensemble Model

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental setup makes use of the CASIA Database, Ubiris, MMU, and Polaris IITD. The Polaris IITD iris dataset consists of 2240 iris images with a resolution of 224 x 64. The Ubiris dataset has 1877 with a resolution of 200 x 150. Around 995 iris images with a resolution of 320 x 238 are considered in the MMU iris dataset and the CASIA iris dataset includes 756 iris images with a 320 x 280 pixels resolution. We use GPUs and the TensorFlow 2.x framework on Google Collaboratory to carry out our experiments. Using a stratified sample technique, we divide the dataset into training, validation, and testing groups. The final features extracted are input to a fully connected layer using the SoftMax activation function. Each convolution layer is related to a rectified linear unit (ReLU) activation function implementing augmentation and the batch normalization technique to reduce the internal covariate shift and overfitting in the activation layer at deeper networks. The dataset is split in such a way that we use 60 percent for training, 20 percent for validation and 20 percent for testing.

The architecture for our neural network is fine-tuned by evaluating the training and validation losses under different hyper parameter settings. The hyper parameter uses the popular optimizer Stochastic Gradient Descent (SGD) with a learning rate of 0.01 and an early stopping technique. It also uses a batch size of 32 layers with 50 epochs, momentum of 0.9, dropout of 50%, label smoothing of 0.1 and a categorical cross entropy as the loss function. The comparative training, validation, testing accuracy of various deep learning models and also the performance of ensemble model based on voting, bagging, stack ensemble and dirichlet ensemble[15][16] for Polaris IITD, Ubiris, MMU and Cassia Dataset are summarized in the Fig 2 to Fig 6. Our experimental research as in Fig 8 shows that an Ensemble-CNN model using DirichletEnsemble accuracy performs better than a single CNN model (Fig 7). The performance of ensemble model is also evaluated using stack ensemble, voting and bagging approach which is represented in the Fig 9 to Fig 11. In Fig 9: the index values in M1234 represent the VGG16, Inception, MobileNet and Nasnet Models respectively. The comparative analysis of the ensemble model with respect to the individual model in terms of performance metrics such as precision, recall and sensitivity is shown in Table 1.

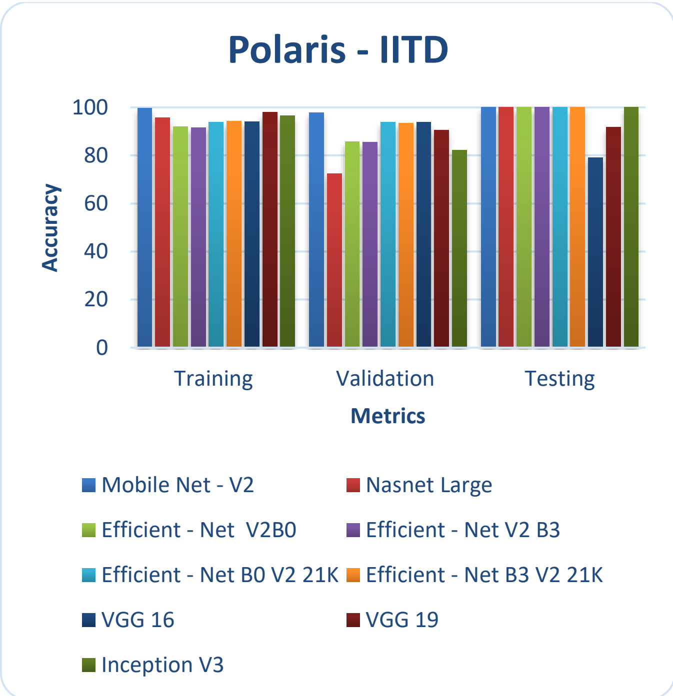


Fig 2: Performance analysis with Polaris Dataset

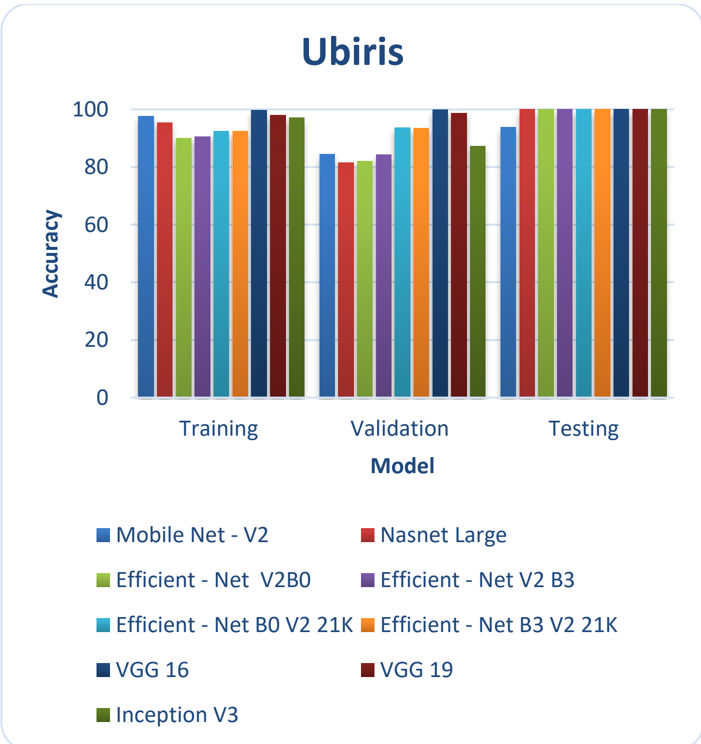


Fig 3: Performance analysis with Ubiris Dataset

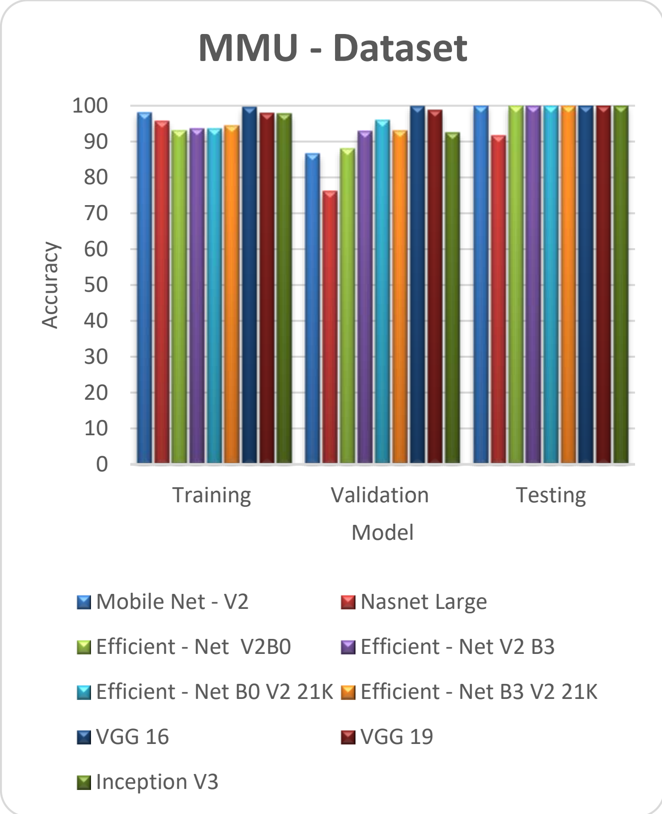


Fig 4: Performance analysis with MMU Dataset

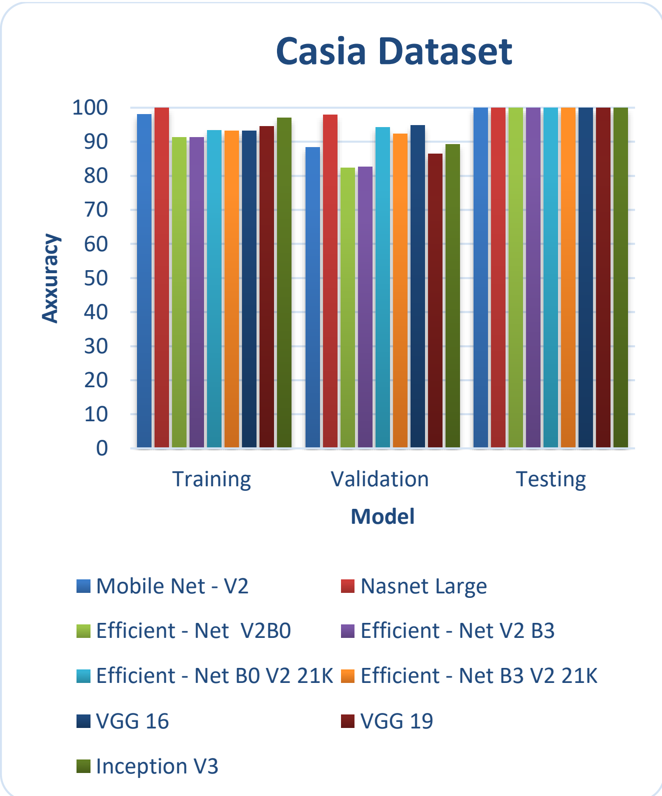


Fig 5: Performance analysis with CASIA Dataset

The performance of traditional texture-based techniques such as principal component analysis, resilient local binary pattern, local ternary pattern, and local binary pattern is also compared with that of deep learning algorithms. Among all the other models used for iris recognition, Nasnet model has the highest recognition accuracy.

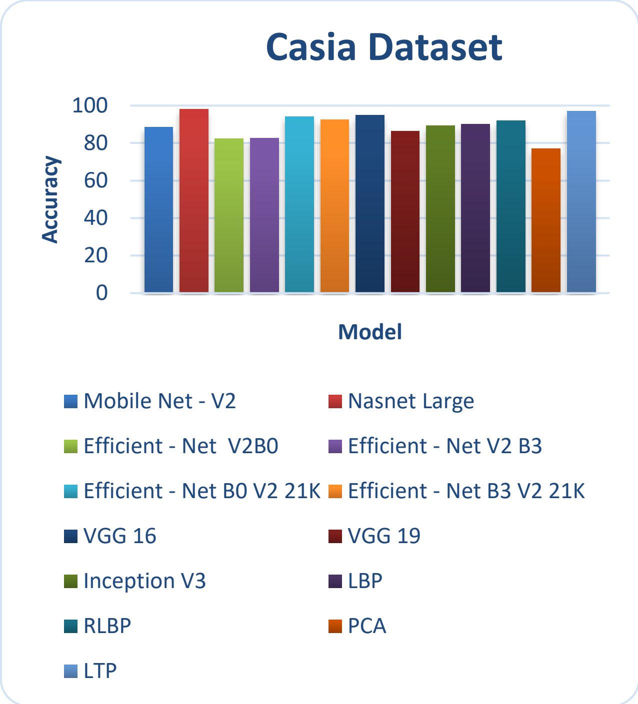


Fig 6: Comparison of DL algorithms with Texture based methods.

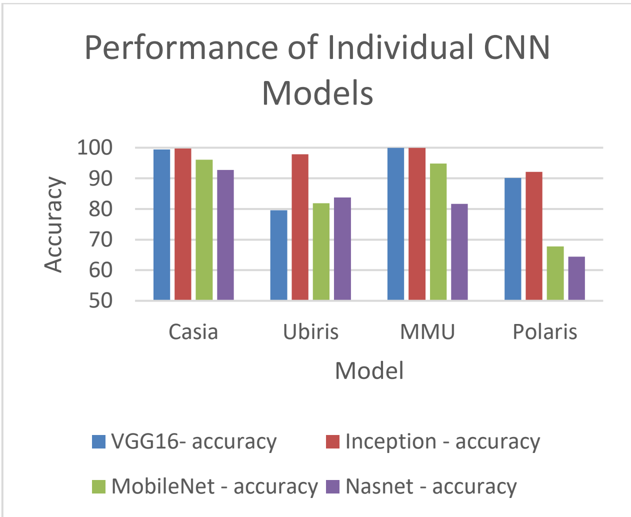


Fig 7: Performance of CNN Models

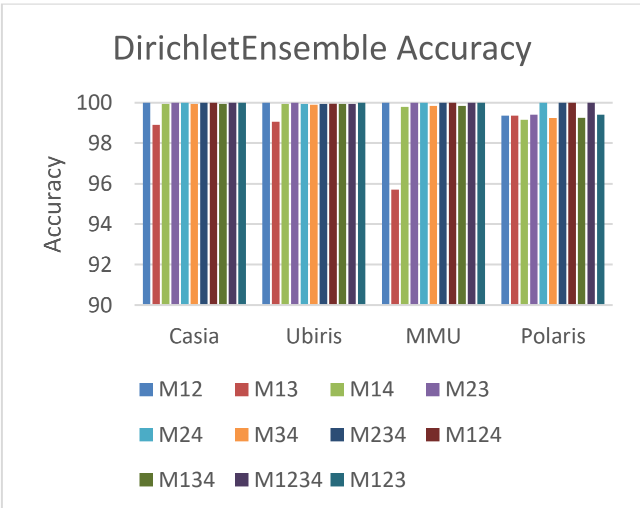


Fig 8: Performance of Ensemble Model due to Dirichlet Ensemble method.

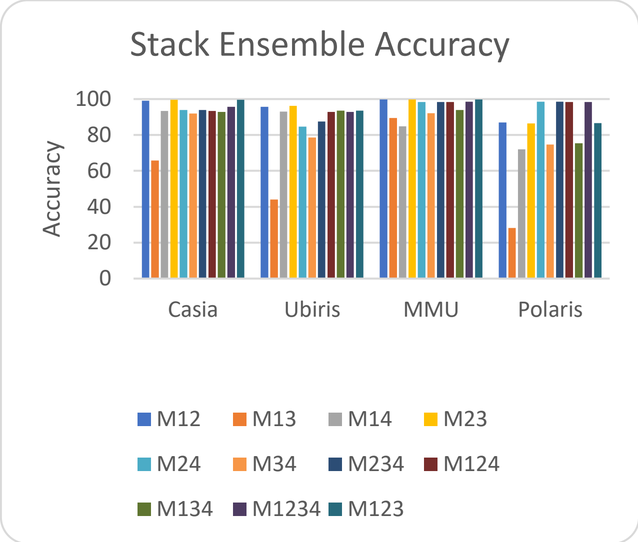


Fig 9: Performance of Ensemble Model using Stack ensemble.

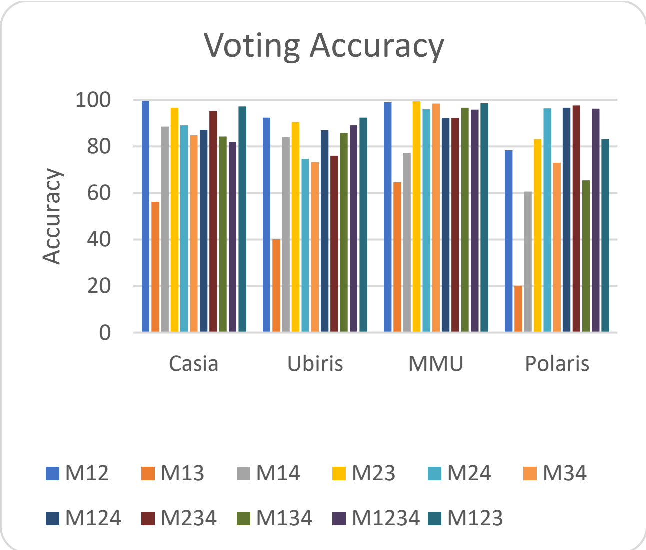


Fig 10: Performance of Ensemble Model using Voting.

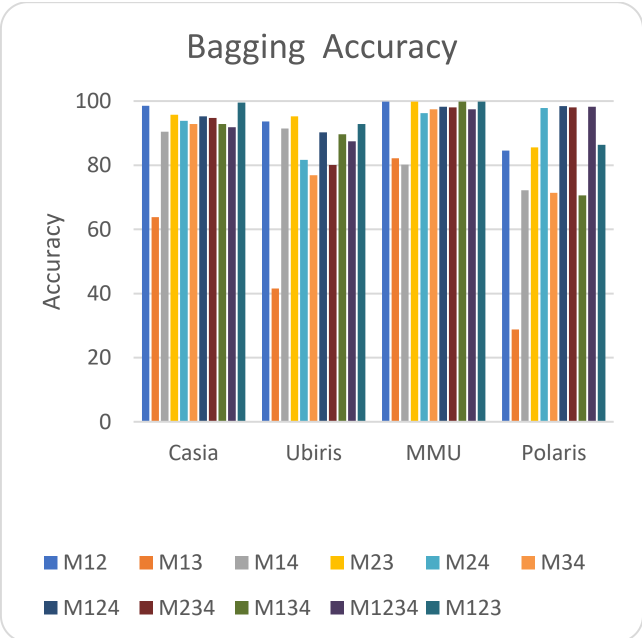
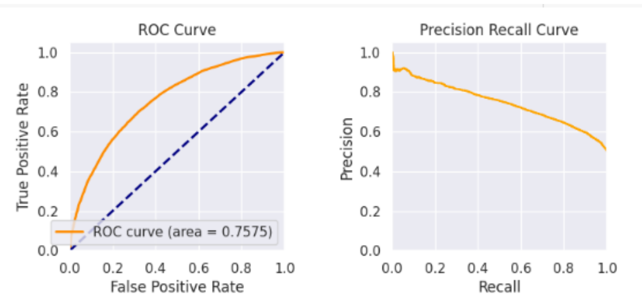


Fig 11: Performance of Ensemble Model using Bagging.

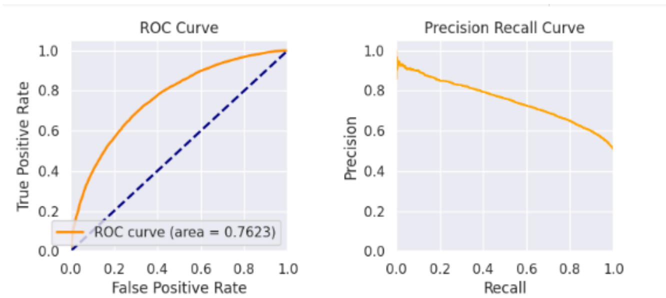
Table 1: Comparative analysis ensemble model and individual models on CASIA, UBIRIS, MMU and Polaris datasets.

Dataset	Model / Metrics	Precision	Recall	F1-Score
CASIA	VGG16	98.1	98.9	98.50
	Inception	97.16	98.24	97.70
	MobileNet	95.24	96.58	95.91
	Nasnet	93.2	93.8	93.50
	Ensemble Model	98.84	100	99.42
UBIRIS	VGG16	96.2	96	96.10
	Inception	83.3	95	88.77
	MobileNet	93.5	97.2	95.31
	Nasnet	65	98	78.16
	Ensemble Model	97.35	98.2	97.77
MMU	VGG16	98.28	93.4	95.78
	Inception	98.45	99	98.72
	MobileNet	92.5	94.7	93.59
	Nasnet	87.1	86.06	86.58
	Ensemble Model	99.35	100	99.67
POLARIS	VGG16	94.66	96.8	95.72
	Inception	92.9	88.7	90.75
	MobileNet	95.32	94	94.66
	Nasnet	83.37	82.2	82.78
	Ensemble Model	96.42	100	98.18

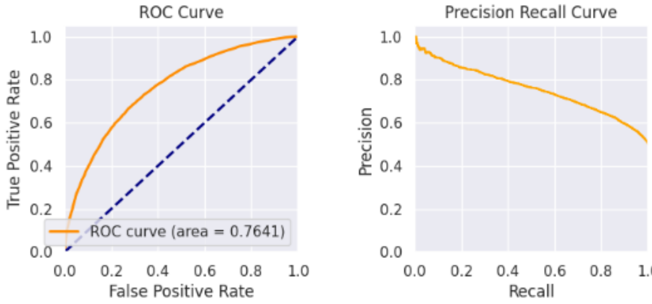
In addition, we have also computed the ROC curve, precision (P) and recall(R) due to the existing deep learning models on different datasets and compared our ensemble method. The results are presented in Fig 12(a-r).



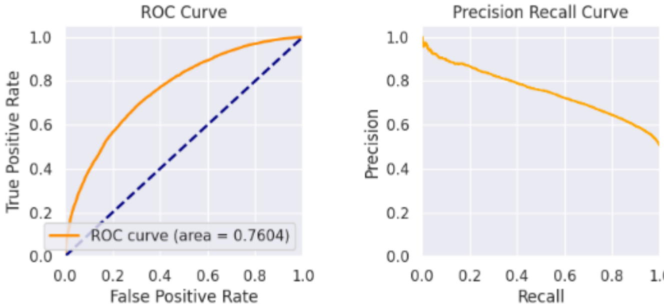
(a) ROC, P and R for Inception model on CASIA dataset.



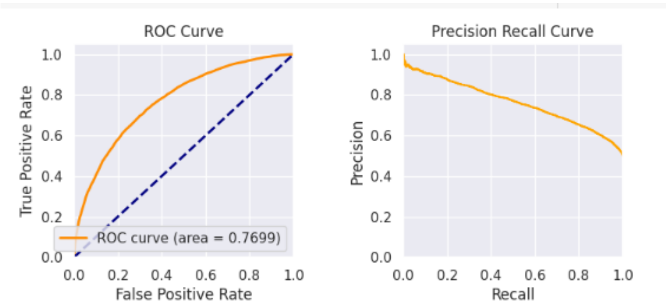
(b) ROC, P and R for VGG16 model on CASIA dataset.



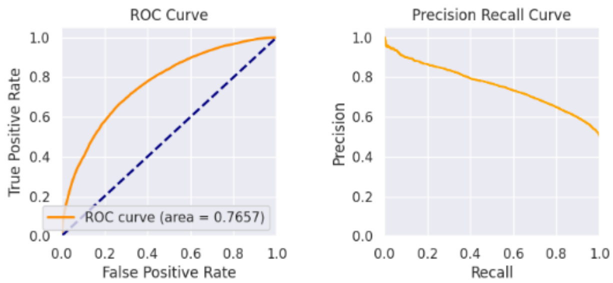
(c) ROC, P and R for Mobilenet model on CASIA dataset.



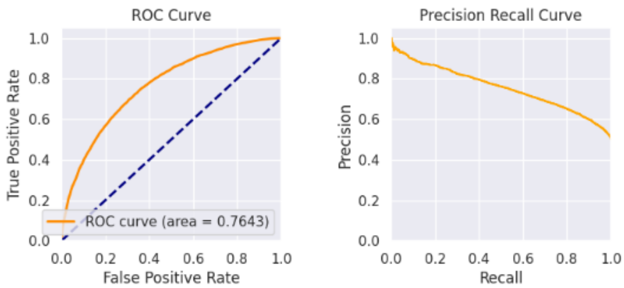
(d) ROC, P and R for Inception model on MMU dataset.



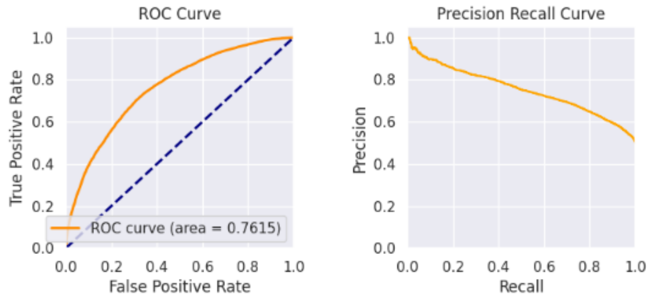
(e) ROC, P and R for VGG16 model on MMU dataset.



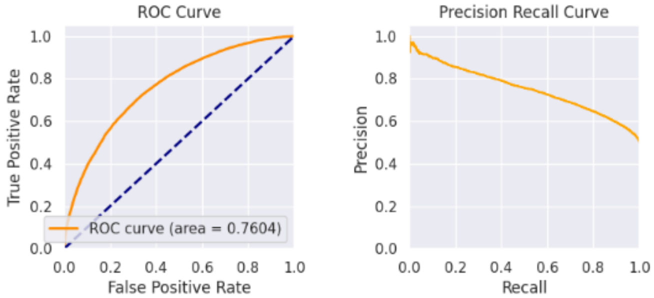
(f) ROC, P and R for Mobilenet model on MMU dataset.



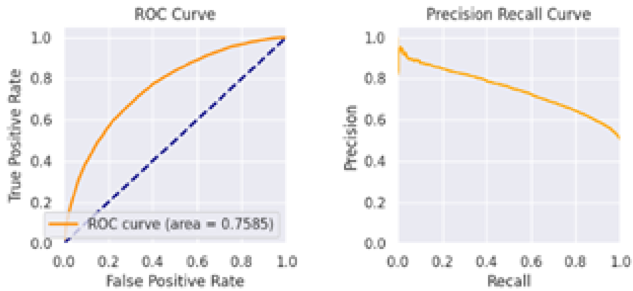
(g) ROC, P and R for Inception model on UBIRIS dataset.



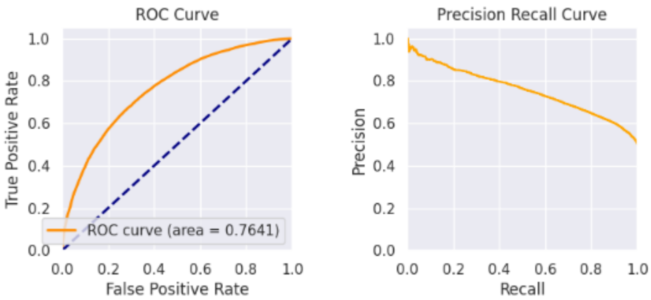
(h)ROC, P and R for VGG16 model on UBIRIS dataset.



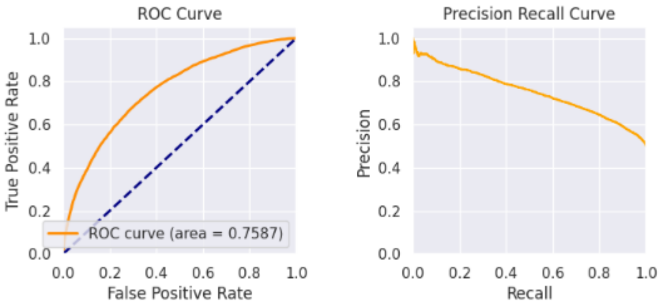
(i) ROC, P and R for Mobilenet model on UBIRIS dataset.



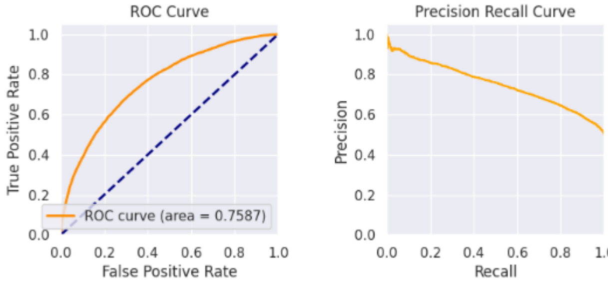
(j) ROC, P and R for NASnet model on UBIRIS dataset.



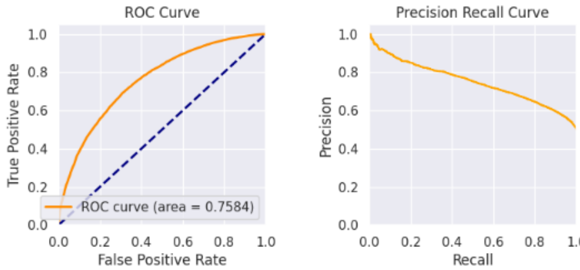
(k)ROC, P and R for Inception model on Polaris dataset.



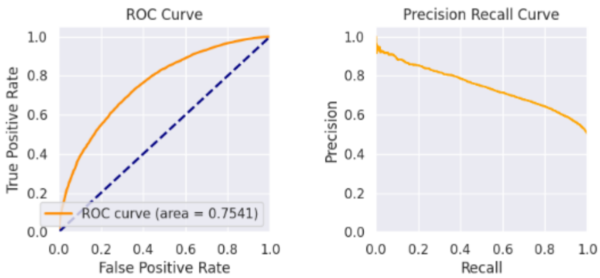
(l) ROC, P and R for VGG16 model on Polaris dataset.



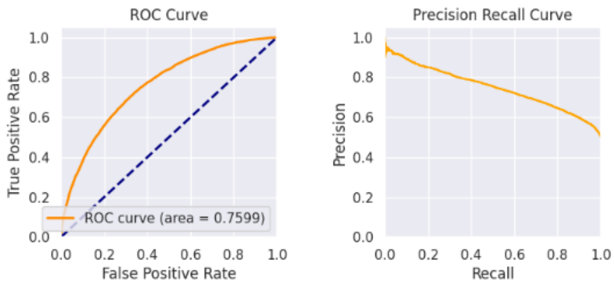
(m) ROC, P and R for Mobilenet model on Polaris dataset.



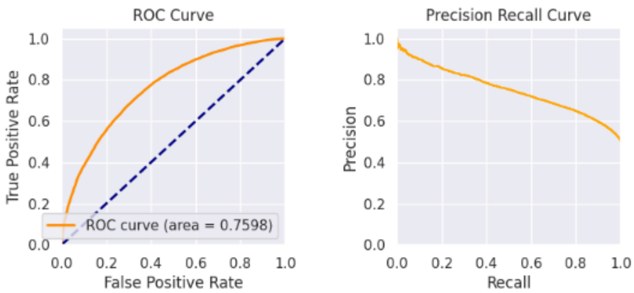
(n)ROC, P and R for NASnet model on Polaris dataset.



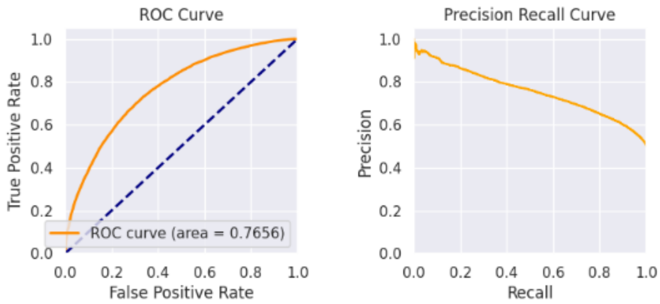
(o)ROC, P and R for **Ensemble model** on CASIA dataset.



(p)ROC, P and R for **Ensemble model** on MMU dataset.



(q)ROC, P and R for **Ensemble model** on UBIRIS dataset.



(r) ROC, P and R for **Ensemble model** on Polaris dataset.

Fig. 12. The ROC, precision (P) and recall (R) values for baseline deep learning models and proposed ensemble model on different iris datasets.

It shall be observed from the above experimental analysis, the proposed ensemble model exhibit better accuracy when compared to the baseline models.

CONCLUSION

In this work, an ensemble based approach for iris biometric identification is presented. It is experimentally observed that the VGG16 perform better on the Ubiris and MMU datasets, while the Nasnet large model does well on the CASIA dataset. The Mobile net model does better on the Polaris IITD dataset. The ensemble model built by considering the best features extracted from each individual model and hence ensure the best performance for a personal identification system.

REFERENCES

1. Anil Jain, Lin Hong, and Sharath Pankanti. Biometric identification. Communications of the ACM, 43(2):90{98, 2000.
2. David Zhang. Automated biometrics: Technologies and systems, volume 7. Springer Science & Business Media,2000.
3. Thakkar, Sejal, and Chirag Patel. "Iris Recognition Supported best Gabor Filters and Deep learning CNN Options." 2020 International Conference on Industry 4.0 Technology (I4Tech). IEEE, 2020.
4. Wang, Kuo, and Ajay Kumar. "Toward more accurate iris recognition using dilated residual features." IEEE Transactions on Information Forensics and Security 14.12 (2019): 3233-3245.
5. Sardar, Mousumi, Subhashis Banerjee, and Sushmita Mitra. "Iris Segmentation Using Interactive Deep Learning." IEEE Access 8 (2020): 219322-219330.
6. Kerrigan, Daniel, et al. "Iris recognition with image segmentation employing retrained off-the-shelf deep neural networks." 2019 International Conference on Biometrics (ICB). IEEE, 2019.
7. Khalifa, Nour Eldeen M., et al. "Deep iris: deep learning for gender classification through iris patterns." Acta Informatica Medica 27.2 (2019): 96.
8. Li, Yung-Hui, Po-Jen Huang, and Yun Juan. "An efficient and robust iris segmentation algorithm using deep learning." Mobile Information Systems 2019 (2019).
9. Minaee, Shervin, and Amirali Abdolrashidi. "Deepiris: Iris recognition using a deep learning approach." arXiv preprint arXiv:1907.09380 (2019).
10. Bazrafkan, Shabab, and Peter Corcoran. "Enhancing iris authentication on handheld devices using deep learning derived segmentation techniques." 2018 IEEE international conference on consumer electronics (ICCE). IEEE, 2018.
11. Zanlorensi, Luiz A., et al. "The impact of preprocessing on deep representations for iris recognition on unconstrained environments." 2018 31st SIBGRAPI Conference on Graphics, Patterns, and Images (SIBGRAPI). IEEE, 2018.
12. Karakaya, Mahmut. "Deep Learning Frameworks for Off-Angle Iris Recognition." 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). IEEE, 2018.
13. Arsalan, Muhammad, et al. "Deep learning-based iris segmentation for iris recognition in visible light environment." Symmetry 9.11 (2017): 263.
14. Hari Narayanan A G, Dr J Amar Pratap Singh "Skin Disease Ensemble Classification Using Transfer Learning and Voting Classifier", International Journal of Engineering Trends and Technology, Volume 69 Issue 12, 287-293, December 2021, ISSN: 2231 – 5381 doi:10.14445/22315381/IJETT-V69I12P234.
15. Hala,Shaaban, " Iris segmentation in uncooperative and unconstrained environments: State of the art datasets and future research directions", Elsevier, Digital Signal processing

118(2021)103244,
https://doi.org/10.1016/j.dsp.2021.103244.
16. M.A.Ganaie, Minghui, Malik, Tanveer, Suganthan, "Ensemble Deep Learning: A Review", Elsevier, Engineering Applications of Artificial intelligence 115(2022)105151. https://doi.org/10/1016/j. engappai. 2022.105151.