https://dx.doi.org/10.4314/dujopas.v10i4b.29
ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 10 No. 4b December 2024*

Development of a Machine Learning Model for Age Prediction of Footballers

Gbenga O. Ogunsanwo*, Blessing C. Okogbue, Godwin O. Odulaja, Ayoade A. Owoade

Department of Computer Science,
College of Science and Information Technology,
Tai Solarin University of Education,
Ogun State,
Nigeria.

Email: ogunsanwogo@tasued.edu.ng

Abstract

This study developed a novel age prediction model that can be adopted in sport especially for football athletes to curb some problems associated with selecting players, and most especially, to minimize age falsification problems. The study aimed at developing an age prediction model applying deep learning with machine learning algorithms. The study acquired age dataset from the FIFA website. The dataset was downloaded in CSV format which contains information about the players, especially their age. Subsequently, we established a database where each player's image was labelled and mapped with their corresponding CSV file. Euclidean Distance (ED) was used for feature reduction. The study employed neural networks with deep learning (DNN) and regression using support vectors (SVR) to develop a model for age prediction. Utilizing Accuracy score, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE), the effectiveness of support vector regression and deep neural network models was assessed. The model was simulated using Anaconda Jupyter IDE. The results showed that the Deep Neural Network model has the MSE value of 49.93, RMSE value of 6.92, MAE value of 5.50 and 81% accuracy, while Support Vector Regression has the MSE value of 35.79, RMSE value of 5.98, MAE value of 5.2 and 82% accuracy. The outcome of the age prediction model developed with the DNN and SVR revealed that the SVR model outperformed DNN. The study recommends that the age prediction model can be used in sport to help the managers in decision making, especially to minimize age falsification problems.

Keywords: Age, Prediction, DNN, SVR, Footballer.

INTRODUCTION

Age identification plays an important part of our social lives. Every language in the world has distinct greetings for men and women, and speaking to seniors frequently requires using a different vocabulary than speaking to younger people. The ability to determine a person's age and gender, which are determined by looking at their face, is a major factor in these practices. (Savchenko, 2019). Particularly considering the expansion of social media and social networks, a great deal of application developers is abusing automatic age detection. The most important aspects of a person's face in social interactions are their age and gender. The identity, age, gender, emotions, and ethnicity can all be determined by a person's traits on their face. In many real-world applications, such as visual surveillance, medical diagnosis

Author for Correspondence

(premature facial aging), human-computer interaction systems, access control or soft biometrics, demographic data collection, law enforcement, and marketing intelligence, age and gender identification can be extremely useful, even sport is not exempted (Dantcheva et al. 2015). In spite of this, the field of computer vision continues to enhance state-of-the-art methods through the introduction of new ones. However, the demands of commercial and practical applications for age and gender predictions from unprocessed real-life facial photos have not yet been satisfied. Therefore, it becomes imperative to have a reliable and precise method for age prediction activities.

The concept of age in relation to sport is a significant factor that can impact an athlete's performance, development, and career trajectory. Age can influence an athlete's physical abilities, psychological attributes, and potential for success in a particular sport (Hancock et al., 2013). Young people are the most important and valuable resource of the nation in different sectors especially in sport because the human body expands and changes rapidly during childhood. The body then takes on its ultimate size and shape as a person approaches their teenage years due to changes in hormone levels like testosterone and estrogen. Under organized training, the human body can reach its peak performance between the ages of 15 and 30. The skeleton becomes less resilient and the body more vulnerable to fractures and other injuries as one approaches the age of thirty, but things start to change gradually after that. The bones start to lose calcium and other minerals over time, resulting in a lower bone density. (Sellami et al. 2021) Simultaneously, the muscles lose an increasing amount of water, calcium, and electrolytes. This causes the muscle mass to gradually start declining, which also causes the body's total strength to decline. As we age, the suppleness of our ligaments and joints gradually diminishes. This limits our range of motion, stiffening the body and causing some activities much harder to execute. (Abate et al., 2010).

This gerontology highlights the impact of physical development changes with age and the concept of relative age effects (RAEs) in sports, and these insights are crucial when considering the issue of age falsification. Age falsification, where athletes misrepresent their age to gain a competitive advantage, disrupts fair competition by exploiting the natural physical advantages associated with different age groups. Given that physical performance typically peaks between 15 and 30 years of age, athletes who falsify their age can manipulate their entry into age categories where their advanced physical development provides an unfair edge.

One notable issue in Nigerian sports is age-falsification, where some athletes, particularly in youth competitions, misrepresent their age to gain a competitive advantage (Ajala, 2018). This unethical practice can distort the integrity of age-group competitions and affect the development of genuine young talents. Age falsification in sports in Nigeria has several negative effects on athletes, the sports community, and the overall integrity of competitions. Some significant impacts of age falsification in Nigeria include unfair advantage and stifling genuine talent (Obilor et al., 2023).

There are numerous works that have used machine learning techniques in image processing and identification for instance, Residual Networks (ResNets) have also been demonstrated to be an excellent way to enhance feature extraction capabilities (Srinivas et.al, 2017).

Odion et al. (2022) Employed a pre-trained deep learning model is in classifying a person’s age group from the sclera (white part) of the human eye. The study used a dataset of 2000 Sclera images collected from 250 individuals of various ages, and Otsu thresholding was used to segment the images using morphological processes.

Khera and Kumar (2022) Focused on age-gender specific prediction model for Parkinson's severity assessment. The study employed computation of clinically relevant features using Vertical Ground Reaction Force (VGRF) data from a total of 165 individuals' database consisting 93 Parkinson's and 72 healthy controls.

Wanga and Davies (2019) Employed the employment of Convolutional Neural Network (CNN) for age and gender prediction.

Peersman et al.(2021) Developed a model to automatically detect the age and gender of the people who wrote the messages on social media.The study developed input data. Thereafter developed tokenization and extraction of the feature. Random Forest Classifier Naive Bayes, Decision tree classifier were used for the classification.

Little or no research has been done on Age prediction using using support vector regression and deep neural network., Also using Euclidean distance to calculate the distance between the eyes. Therefore, this paper aimed to fill this gap by developing a model for age prediction using deep learning algorithms and machine learning techniques which can automatically learn intricate features from complex and structured data.

METHODOLOGY

This section described the method of data acquisition and methodology used in developing the Age predictive model.

Study Techniques

The research approach employed for the study is presented in this section. All procedures that were taken to achieve the desired results are shown in this section. This includes data collection, development of machine learning models, training and testing the datasets, model prediction and assessment. An illustration of the implementation method is provided in Figure 1.

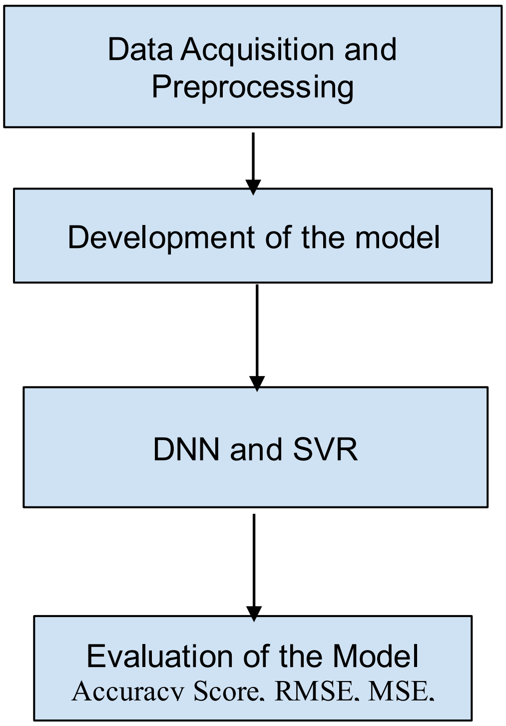


Figure 1 Block Diagram of the Age Predictive Model

Data Acquisition

The initial stage in developing a predictive model is acquisition of relevant dataset. In this research work, the kaggle.com dataset was downloaded in .csv format which contains about 10,000 instances with their various features such as, name, birth date, age, height, weight and nationality as shown in Table 1. Subsequently, a database where each player's image was labelled and mapped with their corresponding CSV file was established. However, to have high accuracy in the predictive model, the dataset that was used for training and validation was subjected to preprocessing.

Table 1: Sample of dataset instances with their features.

full_name	birth_date	age	height_cm	weight_kgs	nationality
Amadou Haidara	31/01/1998	21	175.26	67.1	Mali
Manuel Locatelli	08/01/1998	21	185.42	74.8	Italy
Carles Alena Castillo	05/01/1998	21	154.94	73	Spain
Adrian Czaplak	08/02/1997	22	193.04	88.9	Poland
Benedikt Gimber	19/02/1997	22	185.42	83	Germany
Brian Chevreuil	26/02/1997	22	175.26	78	Haiti
Timo Werner	06/03/1996	23	154.94	74.8	Germany
Marco Asensio Willemsen	21/01/1996	23	182.88	76.2	Spain
Jonathan Tah	11/02/1996	23	195.58	97.1	Germany
Milan Å Skriniar	11/02/1995	24	187.96	79.8	Slovakia
Leon Goretzka	06/02/1995	24	187.96	78.9	Germany
Alessio Romagnoli	12/01/1995	24	187.96	78	Italy
Eric Dier	15/01/1994	25	187.96	89.8	England
Jonathan Castro Otto	03/03/1994	25	170.18	69.9	Spain
Roman Zobnin	11/02/1994	25	182.88	74.8	Russia

Source: www.fifa.com

Data Preprocessing

Gray Scale Conversion of the Coloured Images

In order to improve the dataset's performance of the model, the coloured images were converted into gray scale using the lightness method. The lightness method averages the most prominent and least prominent colours. The pre-processed dataset was subjected to pre training using cascade classifier algorithm.The feature extraction was done using Euclidean Distance (ED) to improve the efficiency of the age predictive model developed. Deep Neural Network Model (DNN Model).The parameters used for the prediction are eye distance, eye width 1, eye width 2, dist_eyes1 to face centre and dist_eyes2 to face centre. Support Vector Regression Model (SVR Model) a subset of Support Vector Machines (SVM). SVR seeks to identify a function that deviates from the real observed values by an amount not to exceed a predetermined margin (epsilon) and at the same time is as flat as possible.

Performance Metrics

The accuracy of the model developed was evaluated using the conventional measurements; MAE, MSE, Accuracy score, and RMSE for each method.

RESULTS AND DISCUSSION

Result of Data Acquisition and Preprocessing

To develop a predictive model for age prediction, the dataset was acquired from FIFA dataset. The FIFA dataset was downloaded in .csv format which contains information about 10,000 players, including their age, but the dataset had to be cleaned and formatted to remove the redundancies and noise. After the data cleaning and formatting, the information of 60 players were retained. Subsequently, a database where each player's image was labelled and mapped with their corresponding CSV file was established. Figure 2 shows the sample of the dataset used for the age prediction model. The dataset was also pre-processed to improve the age predictive model.

Figure 2 Sample of dataset used for the prediction.

Result of Data Preprocessing

In order to improve the dataset's accuracy for the age predictive model, the colored images were converted into gray scale using the lightness method in the gray scale conversion technique. The result of the colored images conversion to gray scale is shown in Figure 3

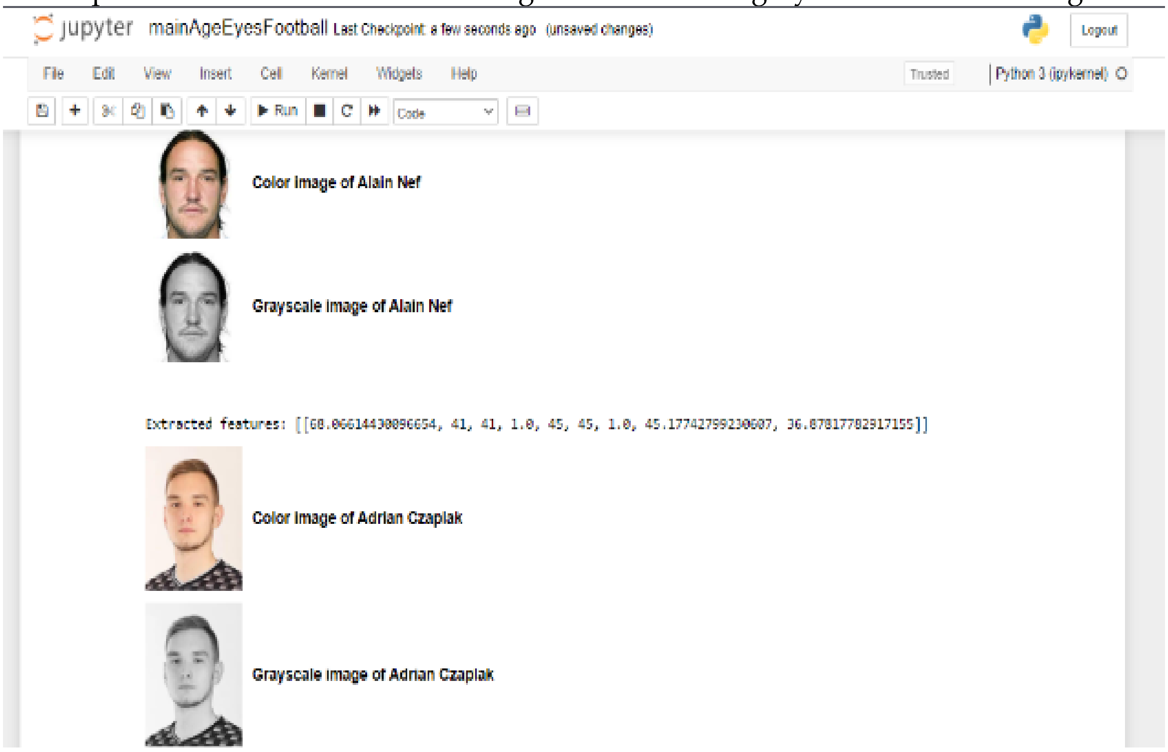


Figure 3 Sample of the gray conversion of the coloured images.

Result of Cascade Classifier Algorithm on Age Prediction Model:

The result of the cascade classifier algorithm used for object detection to encode local intensity patterns in an image for age predictive model is displayed in Figure 4

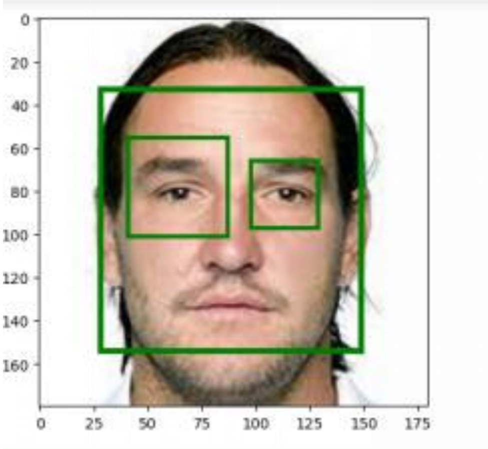
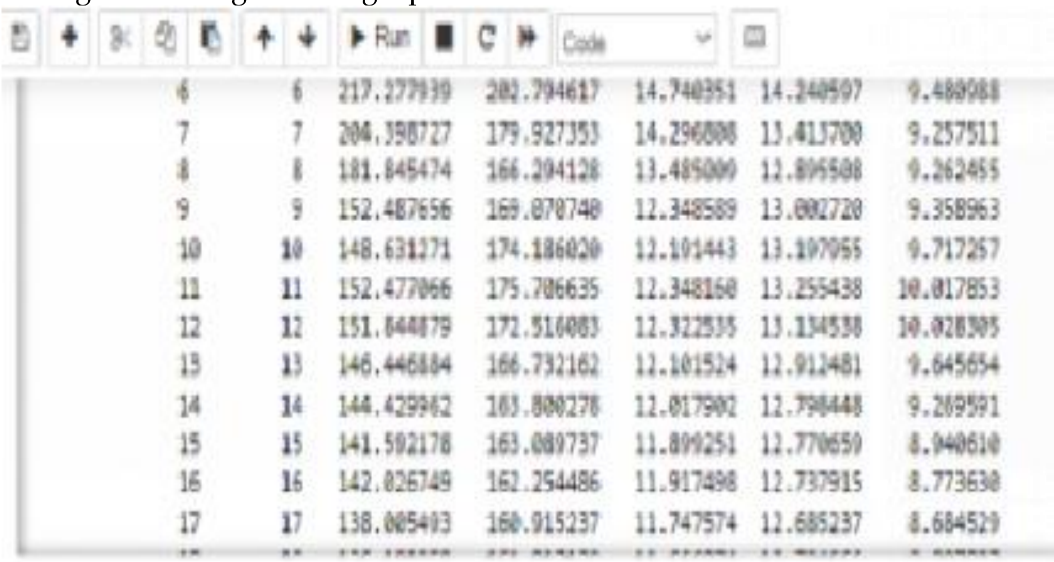


Figure 4 Cascade classifier for Face and Eye detection

Result of Euclidean Distance on Age Prediction Model

The result of Euclidean Distance is shown in Figures 5. The predictive model developed with Euclidean in order to calculate the distance between the parameter used for the prediction. The output shows that each instance's features and corresponding age are properly aligned, allowing for training or testing a predictive model.



6	6	217.277939	202.794617	14.740351	14.240597	9.480988
7	7	204.398727	179.927353	14.296808	13.413700	9.257511
8	8	181.845474	166.204128	13.485000	12.896508	9.262455
9	9	152.487656	169.070740	12.348589	13.002720	9.358963
10	10	148.631271	174.186020	12.101443	13.197065	9.717267
11	11	152.477066	175.706635	12.348160	13.255438	10.017853
12	12	151.844879	172.516083	12.322535	13.134538	10.028305
13	13	146.446884	166.732162	12.101524	12.912481	9.645654
14	14	144.429962	163.800278	12.017902	12.798448	9.269591
15	15	141.502178	163.080737	11.899251	12.770650	8.940610
16	16	142.026749	162.254486	11.917498	12.737915	8.773630
17	17	138.005403	160.915237	11.747574	12.685237	8.684529

Figure 5: A representation of the result of Euclidean distance

DNN Model development

The sample results of the DNN model for age prediction as shown in Figures 6. The result of the output of the DNN model. The code processed a list of image filenames, predicted ages, and actual ages, and then displayed these images with the corresponding predicted and actual ages annotated.

Figure 6 A representation of the results of the DNN model for predicting age

SVR Model development

The sample results of the SVR model for age prediction as shown in Figures 7 and 8. The code processed a list of image filenames, predicted ages, and actual ages, and then displayed these images with the corresponding predicted and actual ages annotated.

```
print(df)
```

	Names	Predicted Age	Actual_Age
0	Adrian Czaplak_1.jpg	31	22
1	Adrian Czaplak_3.jpg	23	22
2	Alain Nef_1.jpeg	37	37
3	Alain Nef_2.jpeg	37	37
4	Alain Nef_3.jpg	30	37
5	Alberto Paloschi_1.png	30	29
6	Alberto Paloschi_3.jpeg	29	29
7	Alessio Romagnoli_3.jpg	25	24
8	Amadou Haidara_3.jpg	22	21
9	Artem Rebrov_1.jpg	35	35
10	Artem Rebrov_2.jpg	32	35
11	Artur Boruc_1.png	31	39
12	Artur Boruc_2.jpeg	32	39
13	Artur Boruc_3.jpeg	30	39
14	Atiba Hutchinson_1.png	36	36
15	Benedikt Gimber_1.jpg	23	22
16	Benedikt Gimber_2.jpg	23	22
17	Benedikt Gimber_3.jpg	23	22

Figure 7 A representation of the results of the SVR model for predicting age

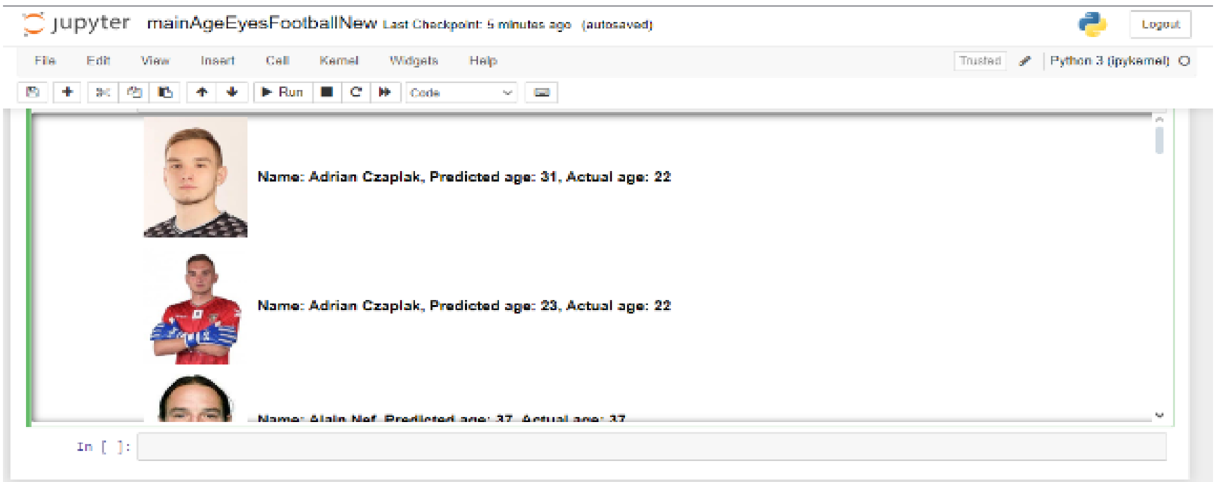


Figure 8 Predicted age with actual age

Result of Model Validation

The metric used for the validation of the age prediction model were MAE, MSE, RMSE, MAPE and Accuracy Score

The Result of MAE

The result of the "Training and Validation Metrics" for the DNN model over a series of epochs. The metrics being plotted was the MAE, for both training as well as dataset validation. Figure 9 displays the Mean Absolute Error (MAE) result for the DNN- based age prediction model. The result clearly revealed that the model learn as the epoch values increases .

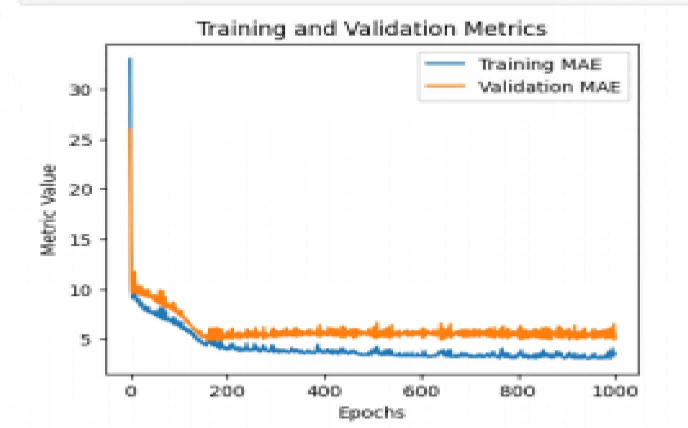


Figure 9 A representation of the result of MAE for the model

The Result of MSE

The result of the "Training and Validation Metrics" for the DNN model over a series of epochs. The metrics being plotted was the MSE, for both training as well as dataset validation. Figure 10 displays the Mean Squared Error (MSE) result for the DNN-based age prediction model. The result clearly revealed that the model learn as the epoch values increases .

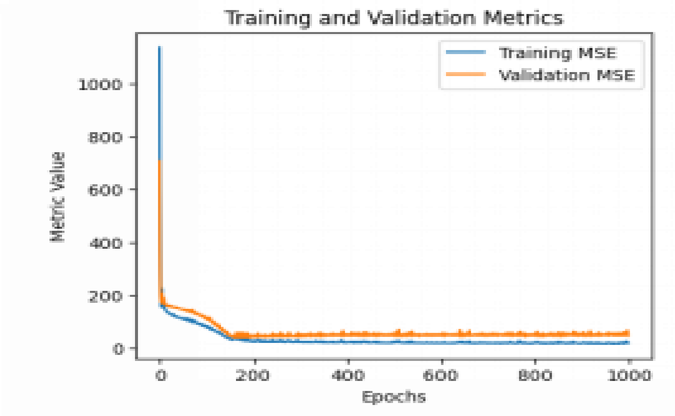


Figure 10 A representation of the result of MSE for the model

The Result of RMSE

The result of the "Training and Validation Metrics" for the DNN model over a series of epochs. The metrics being plotted was the RMSE, for both training as well as dataset validation. Figure 11 displays the Root Mean Squared Error (RMSE) result for the DNN-based age prediction model. The result clearly revealed that the model learn as the epoch values increases .

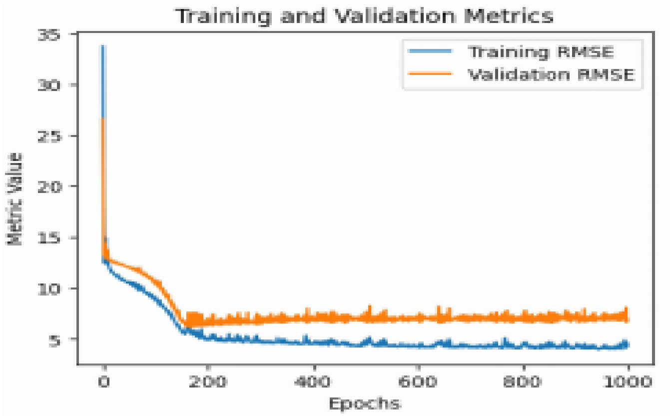


Figure 11 A representation of the result of RMSE for the model

The Result of Accuracy

The accuracy score for the age prediction model using DNN is shown in Figure 12, The "Accuracy Metric" over a series of epochs for the DNN model. The accuracy line is flat and constant at around 81%. This indicates that the accuracy of the model does not change over the epochs and remains stable at 81%.

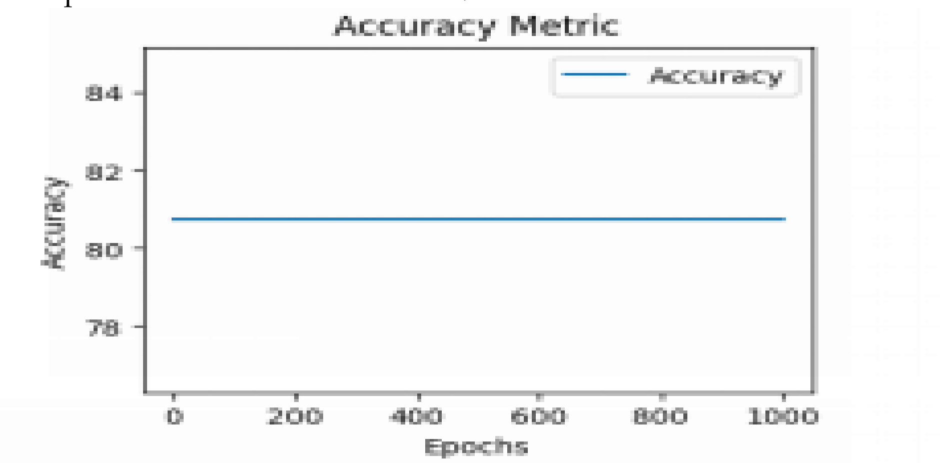
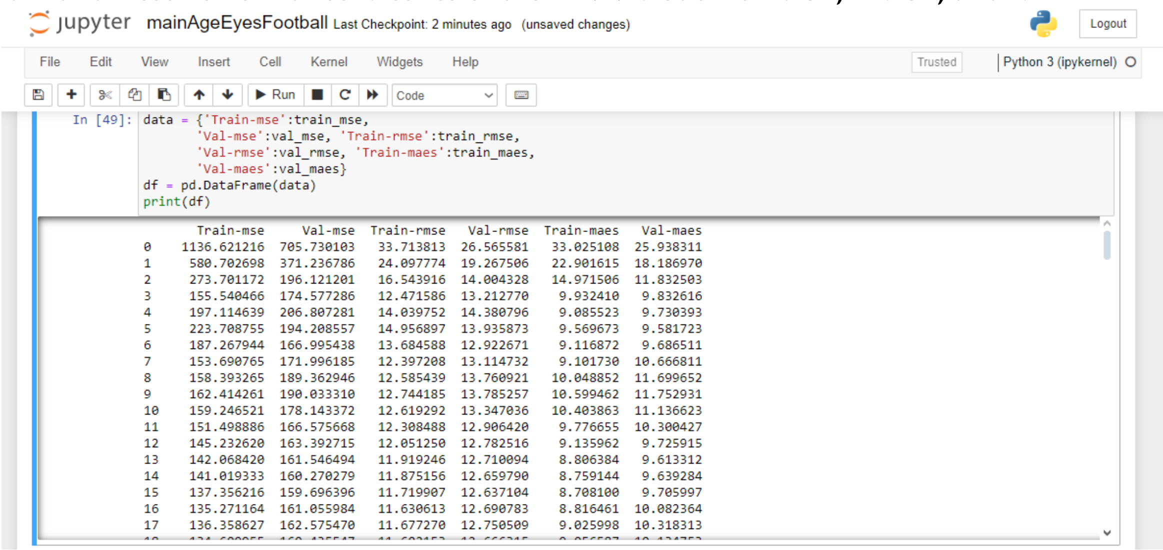


Figure 12 Accuracy score for the age prediction model

Train and Test Performance Metrics of the DNN Model for MSE, RMSE, and MAE




```
data = {'Train-mse':train_mse,
        'Val-mse':val_mse, 'Train-rmse':train_rmse,
        'Val-rmse':val_rmse, 'Train-maes':train_maes,
        'Val-maes':val_maes}
df = pd.DataFrame(data)
print(df)
```

	Train-mse	Val-mse	Train-rmse	Val-rmse	Train-maes	Val-maes
0	1136.621216	705.730103	33.713813	26.565581	33.025108	25.938311
1	580.702698	371.236786	24.097774	19.267506	22.901615	18.186970
2	273.701172	196.121201	16.543916	14.004328	14.971506	11.832503
3	155.540466	174.577286	12.471586	13.212770	9.932410	9.832616
4	197.114639	206.807281	14.039752	14.380796	9.085523	9.730393
5	223.708755	194.208557	14.956897	13.935873	9.569673	9.581723
6	187.267944	166.995438	13.684588	12.922671	9.116872	9.686511
7	153.690765	171.996185	12.397208	13.114732	9.101730	10.666811
8	158.393265	189.362946	12.585439	13.760921	10.048852	11.699652
9	162.414261	190.033310	12.744185	13.785257	10.599462	11.752931
10	159.246521	178.143372	12.619292	13.347036	10.403863	11.136623
11	151.498886	166.575668	12.308488	12.906420	9.776655	10.300427
12	145.232620	163.392715	12.051250	12.782516	9.135962	9.725915
13	142.068420	161.546494	11.919246	12.710094	8.806384	9.613312
14	141.019333	160.270279	11.875156	12.659790	8.759144	9.639284
15	137.356216	159.696396	11.719907	12.637104	8.708100	9.705997
16	135.271164	161.055984	11.630613	12.690783	8.816461	10.082364
17	136.358627	162.575470	11.677270	12.750509	9.025998	10.318313

Figure 13 Trained and validation values of the metrics for the age prediction model (DNN)

Figure 13 shows the result of the train and test performance metrics of the DNN model, specifically for MSE, RMSE, and MAE. The latest data, consistent with previously shown graphs, shows the MSE values for training and testing are 18.81 and 47.93, respectively. 4.33 and 6.92 are the respective RMSE values for training and testing. In addition, the training and testing MAE values are 3.51 and 5.50, correspondingly.

Table 2: Train and Test Performance Metrics of the DNN Model.

Performance Metrics	MSE	RMSE	MAE
Training	18.81	4.33	3.51
Validation	47.93	6.92	5.50

Train and Test Performance Metrics of the SVR Model for MSE, RMSE, MAE and Accuracy

Table 3 shows the result of the train and test performance metrics of the DNN model, specifically for MSE, RMSE, and MAE. It shows the MSE values for training and testing are 0.01 and 35.79, respectively. 0.01 and 5.98 are the respective RMSE values for training and testing. In addition, the training and testing MAE values are 0.01 and 5.2, correspondingly.

Table 3: Train and Test Performance Metrics of the SVR Model.

Performance Metrics	MSE	RMSE	MAE
Training	0.01	0.01	0.01
Validation	35.79	5.98	5.2

DISCUSSION

In this study, DNN and SVR models were were employed in the age prediction regression problem, their results were then compared. The evaluation was conducted using the following metrics: Accuracy Score, RMSE, MSE, and MAE.

The results shows that the Deep Neural Network model has the MSE value of 47.93, RMSE value of 6.92, MAE value of 5.50 and 81% accuracy, while Support Vector Regression has the MSE value of 35.79, RMSE value of 5.98, MAE value of 5.2 and 82% as shown in Table 4

Table 4: Efficiencies of the DNN and SVR models.

Model	MSE	RMSE	MAE	Accuracy Score
DNN	47.93	6.92	5.50	81%
SVR	35.79	5.98	5.2	82%

The Result of MSE

The comparison result of the Mean Squared Error (MSE) for two different models: Deep Neural Network (DNN) and Support Vector Regression (SVR). .The bar graph visually confirms that the MSE for the SVR model is lower than that for the DNN model, indicating that, in this case, the SVR model performs better in terms of prediction as seen in figure 15

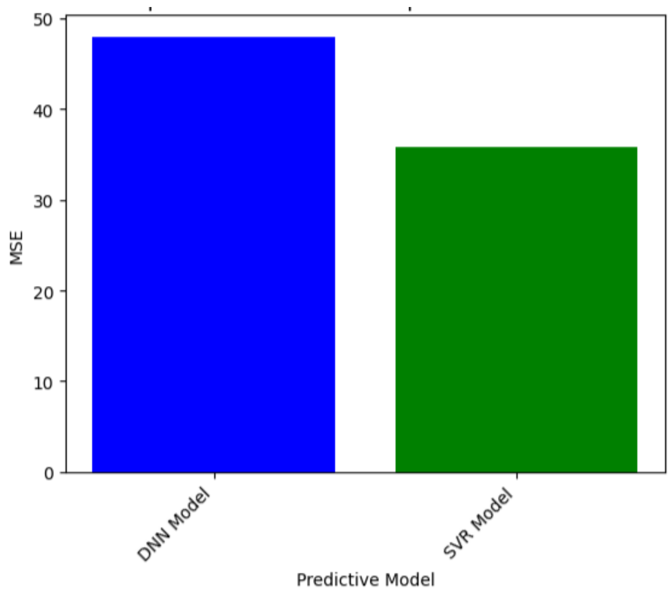


Figure 15 MSE val

The RMSE result

The comparison result of the Root Mean Squared Error (RMSE) for two different models: DNN and SVR. The bar graph visually confirms that the RMSE for the SVR model is lower than that for the DNN model, indicating that, in this case, the SVR model performs better in terms of prediction as seen in Figure 16.

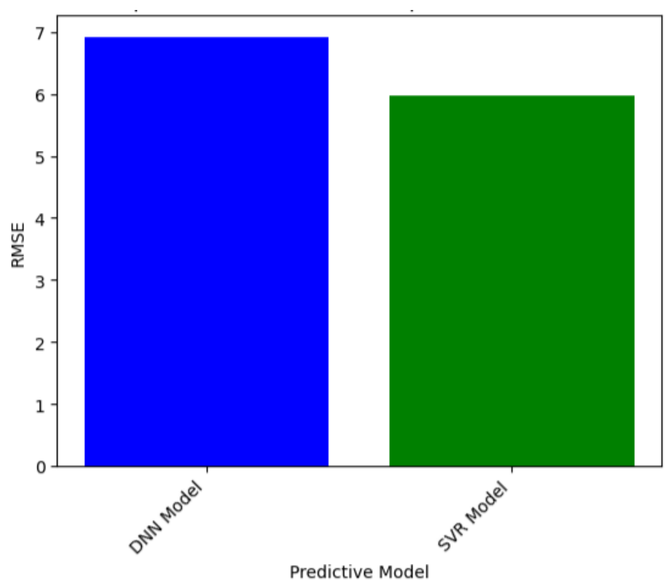


Figure 16 RMSE values to show comparison between DNN and SVR

The MAE Result

The comparison result of the Mean Absolute Error (MAE) for two different models: DNN and SVR..The result confirms that the MAE for the SVR model is lower than that for the DNN model, indicating that, in this case, the SVR model performs better in terms of prediction accuracy as seen in figure 17

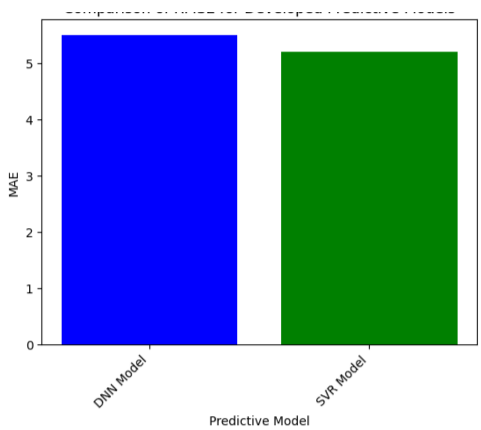


Figure 17 MAE comparison between DNN and SVR

The Accuracy Score

A comparison of the Accuracy Scores for the Deep Neural Network (DNN) and Support Vector Regression (SVR) models is shown in Figure 18.The result of the accuracy shows that-SVR model is higher than the DNN model score, depicting that, in this case, the SVR model performs better in terms of prediction accuracy.

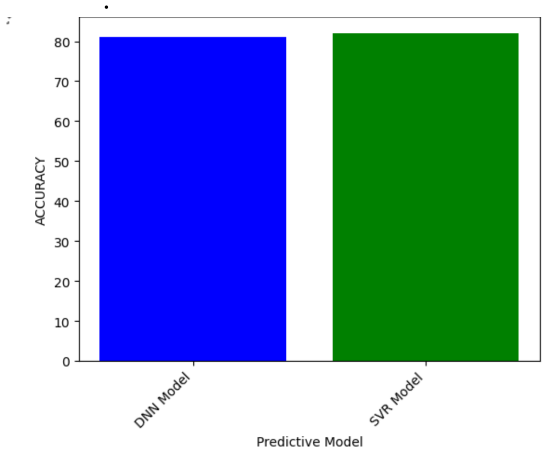


Figure 18 Accuracy score comparison between DNN and SVR

CONCLUSIONS

In this study, machine learning and deep learning were used to predict the footballers age based on their pictures . These predictions are useful in hireing footballers suitable for each age group competition . A rich dataset was obtained from the Kaggle.com. We were able to predict the user's age based on their dataset. The preprocessing of the dataset involved the completion of several tasks, such as clearing up missing values, converting colour images to grey images and uisng EuclIean distnace.. Two major prediction algorithms were adequately implemented in this study, including the SVR, and DNN algorithms. The experiments result of our proposed models indicates that SVR proved to be the more accurate than DNN . Where the detailed analysis of the dataset achieves an accuracy rate of 82% and 81 %. for SVR and DNN respectively Therefore, we can conclude that SVR and DNNs can predict users' ages well. This technique can be employed in Sport to predict users' age based on their pictures type. Although, the obtained results were in line with previous studies. However, our proposed model achieved higher accuracy compare to the study of (Abu et, al. 2020) which achieved 56% accuracy and (Mohammed et, al 2023) which achieved 60% accuracy..The future work can be done using different deep learning techniques and models such as ensemble learning, transfer Learning.

References

Abate, M., Schiavone, C., Pelotti, P., & Salini, V. (2010). Limited joint mobility in diabetes and aging: recent advances in pathogenesis and therapy. *International Journal of immunopathology and pharmacology, 23(4), 997-1003.*

Abu Nada, A.M., Alajrami, E., Al-Saqqa, A.A., AbuNaser, S.S. (2020). Age and gender prediction And validation through single user images using CNN. International Journal of Academic Engineering Research (IJAER), 4(8): 21-24.

Ajala, A. O. (2018). Administrative measures in curbing age-cheating in competitions among football players in Nigeria. *Doctoral dissertation*.

Dantcheva, A., Elia, P., & Ross, A. (2015). What else does your biometric data reveal? A survey on soft biometrics. *IEEE Transactions on Information Forensics and Security, 11(3), 441-467.*

Hancock, D. J., Adler, A. L., & Côté, J. (2013). A proposed theoretical model to explain relative age effects in sport. *European journal of sport science, 13(6), 630-637.*

Khera Preeti & Kumar Neelesh (2022) Age-gender specific prediction model for Parkinson's severity assessment using gait biomarkers Engineering Science and Technology, an International Journal Volume 27, March 2022, 101005

Odiona P. O. Musab M. N & Shuaibua S. U.(2022)Age Prediction from Sclera Images using Deep Learning Journal of the Nigerian Society of Physical Sciences DOI:10.46481/jnsps.2022.787

Obilor, N. M., & Maduagwuna, R. E. (2023). Age Falsification and Work Productivity in *Nnamdi Azikiwe University, Awka. Nnadiebube Journal of Social Sciences, 4(1), 125-135.*

Peersman Claudia DaelemanWalter s & Vaerenbergh Leona Van (2011) Predicting Age and Gender in Online Social Networks ACM 978-1-4503-0949-3/11/10...$10.00.

Savchenko, A. V. (2019). Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet. PeerJ Computer Science, 5, e197.

Sellami, M., Al-Muraikhy, S., Al-Jaber, H., Al-Amri, H., Al-Mansoori, L., Mazloum, N. A., & Elrayess, M. A. (2021). Age and sport intensity-dependent changes in cytokines and telomere *length in elite athletes. Antioxidants, 10(7), 1035.*

Srinivas, N., Atwal, H., Rose, D. C., Mahalingam, G., Ricanek, K., & Bolme, D. S. (2017, May). Age, gender, and fine-grained ethnicity prediction using convolutional neural networks for the East Asian face dataset. *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (pp. 953-960). IEEE.*

Wanga George & Davies Segera Rene(1999) Deep Machine Learning For Age And Gender Prediction ICTACT Journal On Soft Computing, July 2019, Volume: 09, Issue: 04 Doi: 10.21917/Ijsc.2019.0279