Research Article

A deep learning framework for football match prediction

Md. Ashiqur Rahman[1]

Received: 5 August 2019 / Accepted: 28 November 2019 / Published online: 8 January 2020

Abstract

An efficient framework is developed by deep neural networks (DNNs) and artificial neural network (ANNs) for predicting the outcomes of football matches. A dataset is used with the rankings, team performances, all previous international football match results and so on. ANN and DNN are used to explore and process the sporting data to generate prediction value. Datasets are divided into sections for training, validating and testing. By using the proposed DNN architecture, corresponding model performed excellently on predicting the FIFA world cup 2018 matches. This model had predicted 63.3% matches accurately. However, this accuracy can be increased with proper datasets and more accurate information of the teams. The outcome of this hypothesis can be derived that deep learning may be used for successfully predicting the outcomes of football matches or any other sporting events. For more accurate performance of the prediction, prior and more information about each team, player and match is desirable.

Keywords Football match prediction · Deep neural networks · Deep learning · Artificial neural networks

1 Introduction

Football being one of the world's most popular game has a craze in everyone's mind. A football fan must have wanted to know the results before the game at least once in a lifetime! In sport prediction, large numbers of features can be collected including the historical performance of the teams, results of matches, and data on players, to help different stakeholders understand the odds of winning or losing forthcoming matches. The decision of which team is likely to win is important because of the financial assets involved in the betting process; thus bookmakers, fans, and potential bidders are all interested in approximating the odds of a game in advance [1]. The aim of this research is predicting football matches using deep learning algorithms such as ANN and DNN.

Deep neural networks (DNNs) have been used successfully in many scientific, industrial and business domains as a method for extracting knowledge from vast amounts of data. However, the use of DNN techniques in the sporting domain has been limited. Sporting organizations have begun to realize that there is a wealth of untapped knowledge contained in the data and there is great interest in techniques to utilize this data.

The purpose of this study is to develop an efficient framework that will predict football matches correctly and maintain a higher accuracy. An algorithmic study of the prediction techniques is essential for understanding how the deep learning, ANN and DNN technology works to classify item-set.

To predict the outcome of a match between two teams, a person will generally take into account certain factors like the recent performances of the teams, whether the match is going to be played at home or away, recent player transfers, recent coach and staffing changes, etc. The problem when a human predicts the outcome of a match is that the person's decision will be influenced by factors like the human's preference in teams, the human's perception of certain players in a team and so on. Some studies show

✉ Md. Ashiqur Rahman, ashiqur1@cse.wub.edu.bd | [1]Department of Computer Science and Engineering, World University of Bangladesh (WUB), House # 3/A, Road # 4, Dhanmondi, Dhaka 1205, Bangladesh.

SN Applied Sciences (2020) 2:165 | https://doi.org/10.1007/s42452-019-1821-5

decisions may even be influenced by the color of the kit of the team.

The aim of this research is to construct an efficient framework for predicting the result of football matches accurately using deep learning, artificial neural networks and deep neural networks. Figure 1 depicts the LSTM model of this work.

2 Related work

In the last few years, the football betting market has experienced the fastest growth among the gambling markets. Starting with the pioneering works, many econometric methods have been proposed to predict football matches [2, 3]. Tactical pattern recognition in soccer games by means of special self-organizing maps introduced a hierarchical architecture of artificial neural networks to find tactical patterns in soccer game. The hierarchical architecture is capable of recognizing different tactical patterns and variations in these patterns. This study worked on defense player structure using self-organizing map (SOM) as one of the most applicable ANN's technique [4].

Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review has presented a comprehensive literature review on different techniques for CBIR and image representation. This study presents an overview of different techniques that are applied in different research models since the last 12–15 years. After this review, it is summarized that image features representation is done by the use of low-level visual features such as color, texture, spatial layout, and shape [5].

Deeply Learned pose invariant image analysis with applications in 3D face recognition, a novel approach based on deeply learned pose invariant image analysis with applications in 3D face recognition is presented. The PCF alignment algorithm employed the following: (1) pose learning approach using nose tip heuristic to estimate acquisition pose of the face; (2) L2 norm minimization

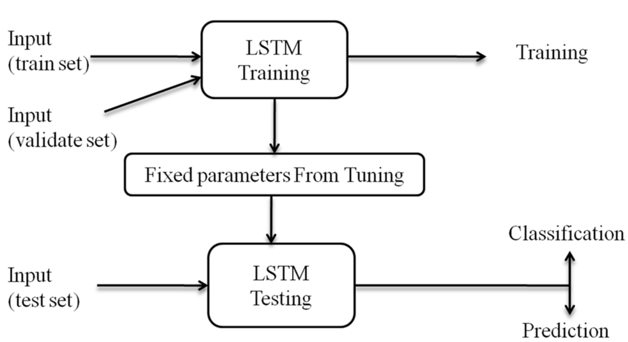


Fig. 1 An illustrative map of the system describing different parts of this study

based coarse to fine approach for nose tip alignment and (3) a transformation step to align the whole face image incorporating the knowledge learned from nosetip alignment [6].

Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach addresses the main challenges of a 3D field deployable face recognition system. It has developed a fully automatic system which uses a composite alignment algorithm to register 3D facial scans with a 3D facial model, thus achieving complete pose-invariance. This system employs a deformable model framework to suit the 3D facial model to the aligned 3D facial scans, and in thus doing measures the distinction between the facial scan and the model in a way that achieves a high degree of expression invariance and thus high accuracy [7].

Data augmentation-assisted makeup-invariant face recognition presented a new data augmentation-assisted makeup-invariant face recognition approach. The planned approach has shown promising results on two standard datasets. Particularly it discussed data augmentation approach based on celebrity-famous makeup styles and semantic preserving transformations suitable for makeup-invariant face recognition. This study focused on training a dCNN model to effectively learn the discriminative features in the presence of cosmetic variations and to fight the frequently occurring problem of overfitting in dCNNs [8].

Modeling global geometric spatial information for rotation invariant classification of satellite image presents a novel approach that computes the spatial clues for the histograms of BoVW model that is robust to the image rotations. The spatial clues are calculated by computing the histograms of orthogonal vectors. This is achieved by calculating the magnitude of orthogonal vectors between pairs of identical visual words (PIVW) relative to the geometric center of an image. The comparative analysis is performed with recently proposed research to obtain the best spatial feature representation for the satellite imagery. This study evaluated the proposed research for image classification using three standard image benchmarks of remote sensing. The results and comparisons conducted to evaluate this analysis show that the planned approach performs better in terms of classification accuracy for a variety of datasets based on satellite images [9].

The impact of asymmetric left and asymmetric right face images on accurate age estimation investigate the role of asymmetric left and asymmetric right face images in accurate age estimation. It has been observed that facial asymmetric aging of two facial halves is associated with estimated age. This study suggests that right asymmetric face images are less inexplicably affected by aging variations compared to the left asymmetric images. This suggests that right asymmetric face images should be used in accurate age estimation [10].

A novel image retrieval based on visual words integration of SIFT and SURF presents a novel visual words integration of scale invariant feature transform (SIFT) and speeded up robust features (SURF).The two local features representations are selected for image retrieval because SIFT is more robust to the change in scale and rotation, while SURF is robust to changes in illumination [11].

2.1 Artificial neural networks (ANN)

Artificial neural networks (ANNs) are a computational approach that is based on the way a biological brain solves problems. ANNs are composed of multiple nodes, which imitate the biological neurons of a human brain. The neurons are connected by links, which imitate the biological axons, and they interact with each other. Like a biological brain, an ANN is self-learning and can therefore excel in areas where the solution is difficult to express by a traditional programming approach. An autoencoder ANN was used in bioinformatics, to predict gene ontology annotations and gene-function relationships [12]. Figure 2 depicts the back-propagation of neural network, where each node takes input data, performs a simple operation, and passes the result to other nodes.

Artificial neural networks (ANNs) consist of a large number of simple units (cells, artificial neurons) working in parallel and exchanging information via a network of directed, weighted links (connections). The information exchanged in these systems usually only comprises the activation level or output of the neurons, a single numerical value for each cell, which is fed to the successor cells after being weighted by their connecting links. In a simplified representation, the topology of an artificial neural network can be seen as a directed, weighted graph with neurons as vertices [13].

2.2 Learning paradigms

Machine learning (ML) is one in every of the intelligent methodologies that have shown promising leads to the domains of classification and prediction. Millilitre specializing in the appliance of artificial neural network (ANN) to sport results prediction. In doing therefore, distinctive the training methodologies utilised, information sources, acceptable suggests that of model analysis, and specific challenges of predicting sport results. One in every of the common machine learning (ML) tasks, that involves predicting a target variable in antecedently unseen information, is classification.

The aim of classification is to predict a target variable (class) by building a classification model supported a coaching dataset, so utilizing that model to predict the worth of the category of check information. This sort of information process is named supervised learning [14]. Unattended ways supported bunch were accustomed distinguish between smart and poor groups. AN ANN with backward-propagation (BP) was then used. Purucker achieved 61% accuracy compared with 72% accuracy of the domain specialists. The BP rule was found to be the foremost effective approach. In reinforcement learning, information x area unit sometimes not given, however generated by AN agent's interactions with the setting.

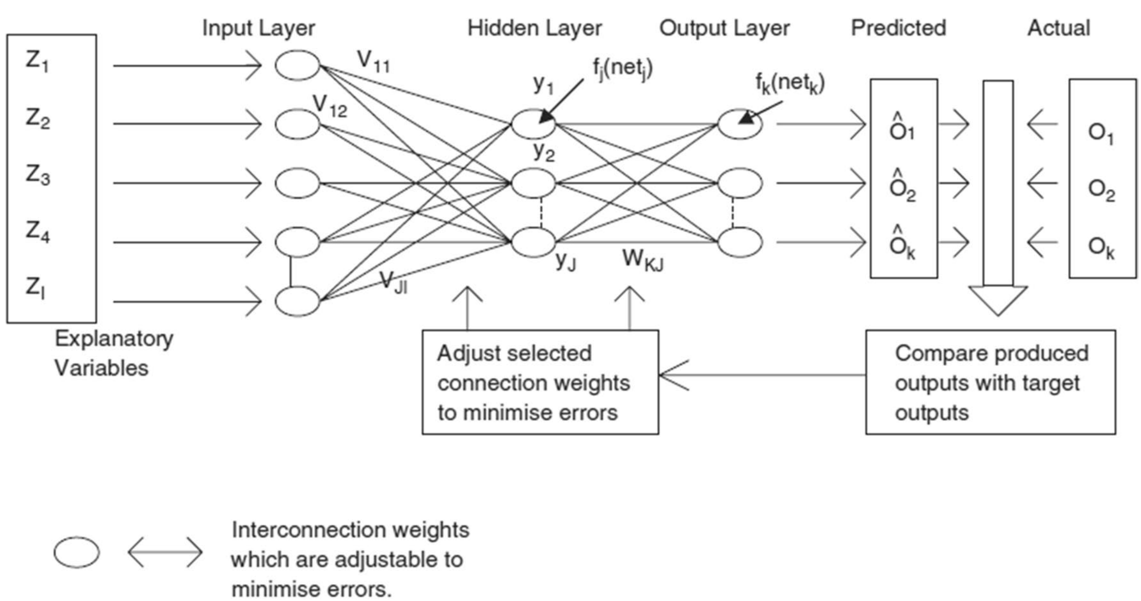


Fig. 2 Basic back-propagation neural network

2.3 Deep neural networks (DNN)

Deep learning is a technique that builds on deep neural networks (DNNs), a form of artificial neural network. The difference is that deep neural networks have multiple hidden layers [15]. There have been big advancements in the field of deep learning during the last few years. Two of the most acclaimed papers are about Deep Mind's network that plays Atari 2600 games and their network beats the world champion in the board game GO. Deep learning-based image recognition has become "superhuman", producing more accurate results than human contestants. This first occurred in 2011 [16].

3 Method

Players representing a soccer club throughout a match are perpetually dynamical. Either because a specific player does not make the team or that a player is transferred to a different team [15]. One of the vital applications in football match prediction that requires good predictive accuracy is match result prediction. Traditionally, the results of the matches are predicted using mathematical and statistical models that are often verified by a domain expert. Due to the specific nature of match-related features to different sports, results across different studies in this application can generally not be compared directly.

The methodology of this study starts with a preprocessing stage of datasets. Preprocessing consists of some different layers. Once dataset sets as input, the model at first reads the dataset and checks the validity of the dataset. If the dataset is not valid or it has some problem, the model may return the dataset.

Figure 3 represent the data flow diagram. Data flow diagram of this work is created after evaluating all the variables and features and analyzing the required processes.

The main focus of this study is to focus on the players initially. In this way it takes into account if a player does not play for a team in a specific match, which will have an impact and it will be recorded. A match is played at a specific time, and events occur at a relative time in a game. The order of matches and events matter since they have an impact on the future. Therefore, a long short-term memory network (LSTM) is exploited for this project.

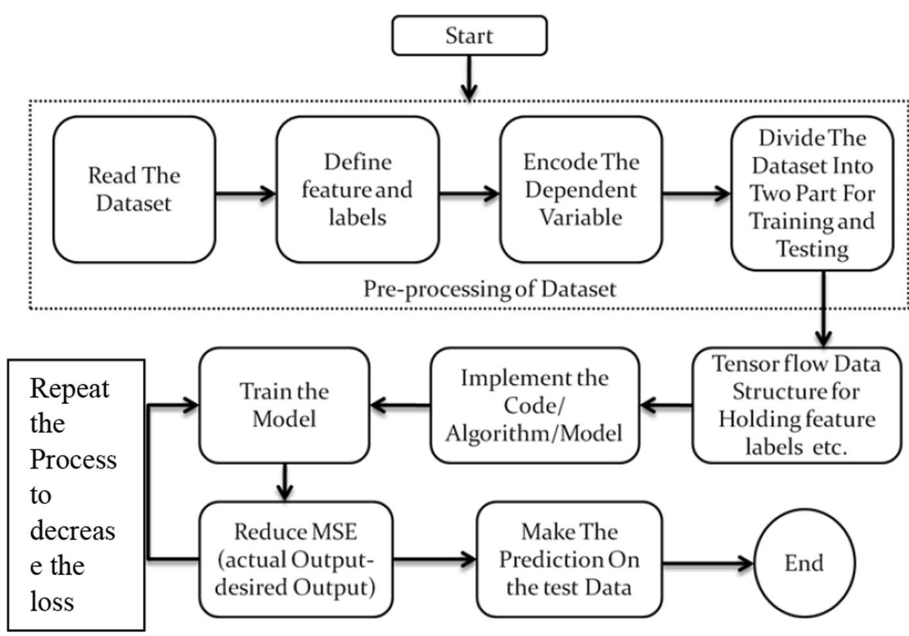
3.1 Deep embeddings

Figure 4 depicts an example of embedded input at time *t* [15].

3.2 Deep learning architecture

Deep learning architectures such as deep neural networks, deep belief networks and recurrent neural networks have been applied to fields including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design and board game programs, where they have produced results comparable to and in some cases superior to human experts [17]. Deep learning models are vaguely inspired by information processing and communication patterns in biological nervous systems yet have various differences from

Fig. 3 Data flow diagram of methodology



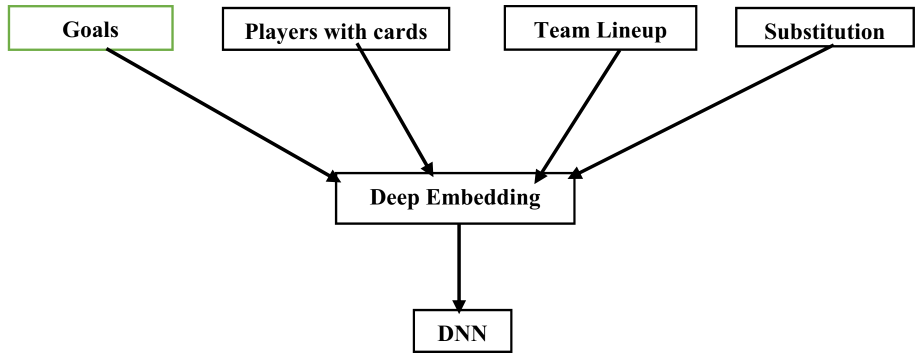


Fig. 4 Deep embeddings of DNN

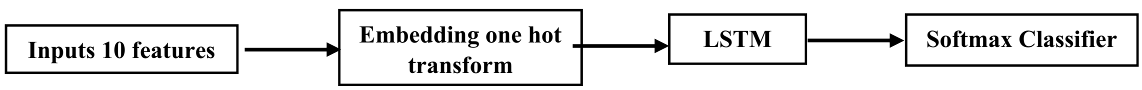


Fig. 5 The simple high level architecture of the model

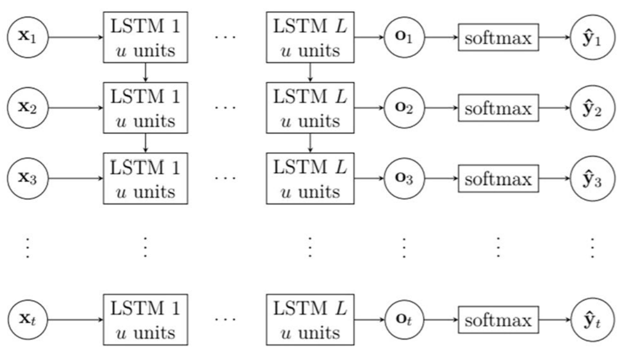


Fig. 6 Detailed illustration of the architecture used

the structural and functional properties of biological brains, which make them incompatible with neuroscience evidences [18].

A few different models have been used and tested in this project. The core DNN consists of LSTM or GRU cells and a softmax classifier

The block diagram of the model can be seen in Fig. 5 that consists of inputs, input processing, some number of LSTM layers and units, and lastly a softmax classifier. This architecture can also be seen in detail in Fig. 6. The input vector consists of 10 number features: amount, home team, away team, main player, aiding player, position, goal type, card type, penalty kind, and substitute.

$\boldsymbol{x_t}$ is the input vector at time step $\boldsymbol{t}$, $\boldsymbol{o_t}$ is a single layer feed forward neural network, $\boldsymbol{y_t}$ is the predicted class at time $\boldsymbol{t}$, $\boldsymbol{L}$ represents the number of layers, and $\boldsymbol{u}$ number of LSTM units per layer.

3.3 Many-to-one

The calculations for accuracy of this technique is described as follows:

For each match $\boldsymbol{m}$, there is a target (i.e., class label) $\boldsymbol{y_m}$:

$$\boldsymbol{y_m} = \begin{cases} [1,0,0] & (\boldsymbol{Home\ Victory}) \\ [0,1,0] & (\boldsymbol{Draw}) \\ [0,0,1] & (\boldsymbol{Away\ Victory}) \end{cases} \tag{1}$$

and an output $\boldsymbol{y_m}$ from the last event in match $\boldsymbol{m}$

$$\boldsymbol{y_m} \in \{[1,0,0],[0,1,0],[0,0,1]\} \tag{2}$$

the accuracy is then calculated by

$$\boldsymbol{r_m} = \begin{cases} 1 & \boldsymbol{if\ y_m = y_m} \\ 0 & \boldsymbol{otherwise} \end{cases} \tag{3}$$

$$\mathbf{accuracy} = \frac{\sum_{m=1}^{M} \boldsymbol{r_m}}{\boldsymbol{M}} \tag{4}$$

where M is the number of matches.

3.4 Many-to-many

The calculations for the accuracy of this technique is described as follows:

For each match $\boldsymbol{m}$, there is a target (i.e., class label) $\boldsymbol{y_m}$:

$$\boldsymbol{y_m} = \begin{cases} [1,0,0] & (\boldsymbol{Home}\ \mathrm{Victory}) \\ [0,1,0] & (\boldsymbol{Draw}) \\ [0,0,1] & (\boldsymbol{Away\ Victory}) \end{cases} \tag{5}$$

$$\boldsymbol{y_m} = \left[\boldsymbol{y}^1_{\boldsymbol{m}}, \boldsymbol{y}^2_{\boldsymbol{m}}, \ldots, \boldsymbol{y}^{\boldsymbol{n}}_{\boldsymbol{m}}\right] \tag{6}$$

where $\boldsymbol{y}^{\boldsymbol{i}}_{\boldsymbol{m}}$ is the predicted outcome for event $\boldsymbol{i}$, and $\boldsymbol{n}$ is the number of events in the match $\boldsymbol{m}$.

The number of correct predictions for each match is calculated by

$$\boldsymbol{r_m} = \begin{cases} 1 & \boldsymbol{if\ y_m = y_m} \\ 0 & \boldsymbol{otherwise} \end{cases} \quad \textbf{for i} = 1, \ldots, \textbf{n} \tag{7}$$

and then the accuracy over all matches is calculated by -

$$\boldsymbol{accuracy} = \frac{\sum_{m=1}^{M} \sum_{i=1}^{n} \boldsymbol{r}^{\boldsymbol{i}}_{\boldsymbol{m}}}{\sum_{m=1}^{M} \#\ \boldsymbol{elements\ in\ r_m}} \tag{8}$$

where M is the number of matches.

3.5 Stochastic gradient descent

Stochastic gradient descent AN intuitive thanks to consider gradient descent is to imagine the trail of a watercourse originating from high of a mountain. The goal of gradient descent is precisely what the stream strives to realize particularly, reach all-time low most purpose (at the foothill) mounting down from the mountain.

3.6 Learning rate decay

This has the result of quickly learning smart weights early and fine calibration them later. Two common and straightforward to use learning rate decay area unit as follows:

- Decrease the learning rate gradually based on the epoch.
- Decrease the learning rate using punctuated large drops at specific epochs.

3.7 Long short term memory (LSTM)

The beauty of the LSTM is that it decides all this supported this input itself as shown within the Fig. 7

The signal x(t) at this time stamp decides all the on top of three points. The input gate takes a call for purpose one. The forget gate takes a call on purpose two and therefore the output gate takes a call on purpose three. This can be galvanized by however our brains work and might handle explosive context switches supported the input.

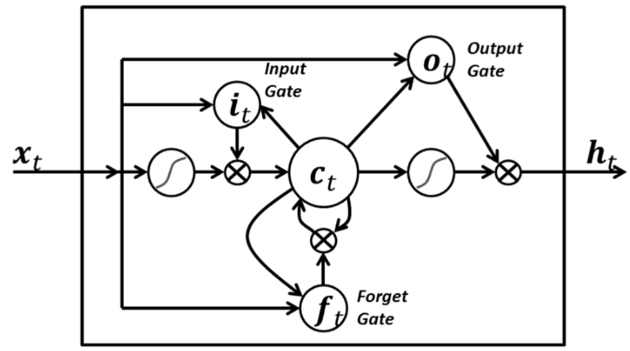


Fig. 7 Long short term memory

4 Results and findings

4.1 Dataset

The dataset used in this research for testing and training is "International Football results from 1872 to 2018", "FIFA world cup 2018 Dataset" and "FIFA Soccer Rankings". This contains collecting relevant columns and adjusting according to other dataset. The validation and testing dataset are normalized to contain roughly the same number of classes, so that the results on the testing set should not be biased.

4.2 Training

Training was a big and time consuming part of this research. Some techniques result in better classification but lower prediction than others. Software tools e.g., NVidia driver 411.63, Python 3.7.0, Tensor Flow 1.11, Anaconda 5.2 are used in the process for training.

Extracted features from the dataset as depicts in Table 1 which also indicate the following attributes:

- Point and rank differences.
- If the game was for some stakes, because naive view was that typically friendly matches are harder to predict.
- How many days the different teams were able to rest but this turned out to be not important enough to be worth the hassle.
- Include the participant countries as a one hot vector but that did not appear to be a strong predictor either.

This Table 2 shows the joining matches with rankings.

Table 1 Features from dataset

Rank	Country full	Country abrv	Total points	Previous points	Rank change	Cur year avg	Cur year avg weighted	Last year avg	Last year_avg weighted	Two-year ago_avg	Two-year ago weighted	Three-year ago avg	Three-year ago weighted	Confederation	Rank date
1	Germany	GER	0	57	0	0	0	0	0	0	0	0	0	UEFA	8/8/1993
2	Italy	ITA	0	57	0	0	0	0	0	0	0	0	0	UEFA	8/8/1993
3	Switzerland	SUI	0	50	9	0	0	0	0	0	0	0	0	UEFA	8/8/1993
4	Sweden	SWE	0	55	0	0	0	0	0	0	0	0	0	UEFA	8/8/1993
5	Argentina	ARG	0	51	5	0	0	0	0	0	0	0	0	CONMEBOL	8/8/1993
6	Republic of Ireland	IRL	0	54	0	0	0	0	0	0	0	0	0	UEFA	8/8/1993
7	Russia	RUS	0	52	1	0	0	0	0	0	0	0	0	UEFA	8/8/1993

Table 2 Joining matches with rankings-

	Date	Home team	Away team	Home score	Away score	Tournament	City	country	Neutral	Rank date home	three_year_ago_weighted_home
0	1993-08-08	Bolivia	Uruguay	3	1	FIFA World Cup qualification	La Paz	Bolivia	False	1993-08-08	0.0
1	1993-08-08	Brazil	Mexico	1	1	Friendly	Maceio	Brazil	False	1993-08-08	0.0
2	1993-08-08	Ecuador	Venezuela	5	0	FIFA World Cup qualification	Quito	Ecuador	False	1993-08-08	0.0
3	1993-08-08	Guinea	Sierra Leone	1	0	Friendly	Conakry	Guinea	False	1993-08-08	0.0
4	1993-08-08	Paraguay	Argentina	1	3	FIFA World Cup qualification	Asuncion	Paraguay	False	1993-08-08	0.0

4.3 Tuning parameters

Following section describes the parameters of LSTM architectures and how they were tuned.

Dropout	The first networks trained did not utilize dropout
Learning rate	Around 10 − 4 to work very well, both in times of convergence and time used to train which is why is chosen to fix it as 10 − 4
Batch size	Batch sizes between 1 and 500 and noticed that sizes over 100 were too large for data and setup
Embedding dimensions	Since there are many different values for the team and players; to continue with a dimension size of 30, and for the rest of the values used was 10 to ease the size of the input for computational reasons [15]

4.4 Classification

Tests of this work are divided into two parts, classification and prediction. Classification uses all the data available and simply does a classification of which class a match belongs to. There are three different classes, home win, draw, and away win.

A Deep Neural Network performs well on classification in football match prediction. Deep network also used in image classification as search for a relevant image from an archive is a challenging research problem for computer vision research community. There are lots of model for image classification.

In CBIR and image classification-based models, high-level image visuals are represented in the form of feature vectors that consists of numerical values. A pose invariant deeply learned Multiview 3D face recognition approach is used to address two problems: face alignment and face recognition through identification and verification setups. Full automation is provided through the use of advanced multistage alignment algorithms, resilience to facial expressions by employing a deformable model framework, and invariance to 3D capture devices through suitable preprocessing steps. Scalability in both time and space is achieved by converting 3D facial scans into compact metadata. The Bag of Visual Word (BoVW) model has been used for the classification of satellite images. In BoVW model, an orderless histogram of visual words without any spatial information is used as image signature. To cope with these artificial effects, a deep convolutional neural network (dCNN) using augmented face dataset to extract discriminative features from face images containing synthetic makeup variations. The augmented dataset containing original face images and those with synthetic makeup variations allows dCNN to learn face features in a variety of facial makeup. The asymmetric aging of the left and right sides of face images and its impact on accurate age estimation. Left symmetric faces were perceived as younger while right symmetric faces were perceived as older when presented to the state-of-the-art age estimator. These findings show that facial aging is an asymmetric process which plays role in accurate facial age estimation. SIFT and SURF, the two local features representations are selected for image retrieval because SIFT is more robust to the change in scale and rotation, while SURF is robust to changes in illumination. The visual words integration of SIFT and SURF adds the robustness of both features to image retrieval.

4.5 Prediction

This section describes the results from prediction. Prediction is performed by using only part of the previous data. The network predicts the future values which have not yet been seen. It created a deep neural network which contains of 2 hidden layers of 10 nodes and 5 nodes. That means this model contains 15 nodes for predicting the matches. It tried different parameters in batch size and learning rate, maximum AUC (Area under the curve) and accuracy found around 0.7. 60% of the data were used for training the model and rest of the 40% were used for validation and prediction.

A comparison between Log Loss and periods is shown in the Fig. 8. Log loss measures the performance of a classification model where the prediction input is a probability value between 0 and 1. The goal of the machine learning models is to minimize this value. A perfect

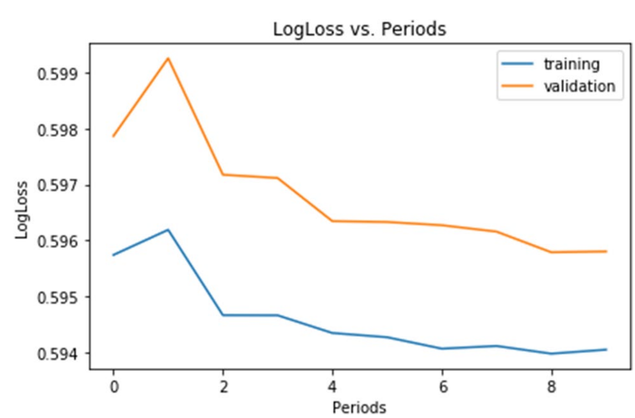


Fig. 8 Checking LogLoss versus Periods

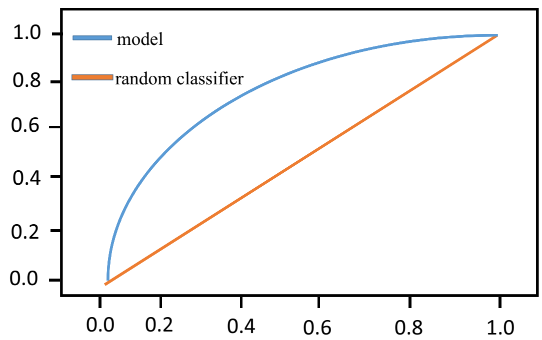


Fig. 9 Checking the AUC and accuracy of validation set

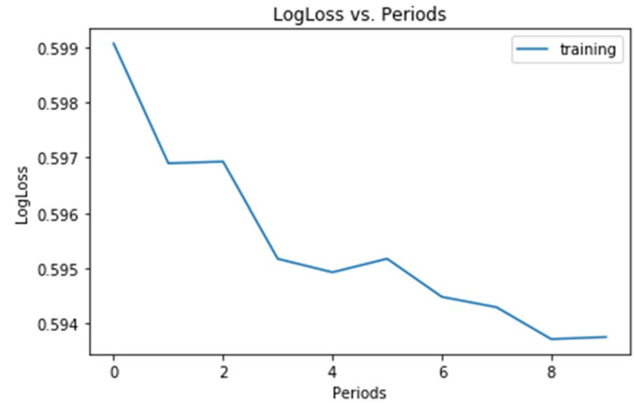


Fig. 10 LogLoss versus Periods

model would have a log loss of 0. Log loss increases as the predicted probability diverges from the actual label.

4.5.1 Checking AUC and accuracy of DNN classifier model

After evaluating the AUC (Area under the curve) and accuracy during testing, the model had an AUC of 0.74 on the validation set and accuracy of 0.67 on the validation set. A Fig. 9 showing the accuracy and area under the curve (AUC) of the validation set is given below -

4.5.2 Training model

A demo of training period is given below

LogLoss (on training data):

period 00: 0.60
period 01: 0.60
period 02: 0.59
period 03: 0.59
period 04: 0.59
period 05: 0.59
period 06: 0.59
period 07: 0.59
period 08: 0.59
period 09: 0.59

Model training finished

After the model is trained, the data is converted into a dict of np arrays. That constructs a dataset and configures batching, repeating. The data can be shuffled if user specifies it. After completion of one batch, return the next batch of data. There is a DNN regressor object created and also created some input functions for training the model (Fig. 10).

4.5.3 Predicted result versus actual result (FIFA World Cup 2018)

While trying to simulate FIFA World cup 2018, a small margin is used where it is safer to predict to draw then win and then also defined the team rankings at the time of world cup. After the completion of training part, the model is trained with world cup data and tried to predict FIFA World Cup 2018 matches. Here, a comparison between the actual result and predicted result is demonstrated in the following two Tables 3 and 4.

A comparison between the actual result and predicted result of group A, B, C, D.

Here, a comparison between the actual result and predicted result of group E, F, G, H.

5 Conclusion

A Deep Neural Network performs well on classification and has potential to perform well on predictions of the football match results. Using the proposed DNN architecture, this model performed excellently on predicting the FIFA world cup 2018 matches. This model had predicted 63.3% matches accurately of FIFA World cup 2018 group stage matches. Different datasets can be used, different inputs and architectures can be tested and other things than winner can be predicted, such as the amount of goals or cards including in which minute a team will score. Different sport data can be used for different predictions. Different datasets may lead to predict other sport field matches like cricket, hockey etc.

5.1 Limitations

Though this research is successfully completed, this work had some limitations to deal with. Some of the major limitations were –

Table 3 Predicted result versus actual result

A			B			C			D		
Team name	Actual Result	Predicted Result	Team name	Actual Result	Predicted Result	Team name	Actual Result	Predicted Result	Team name	Actual Result	Predicted Result
Russia versus Saudi Arabia	Russia 5–0	Saudi Arabia wins with 0.67	Morocco versus Iran	Iran 1–0	Iran wins with 0.76	France versus Australia	France 2–1	France wins with 0.60	Argentina versus Croatia	Croatia 3–0	Draw
Russia versus Egypt	Russia 3–1	Egypt wins with 0.67	Portugal versus Spain	3–3	Draw	Peru versus Denmark	Denmark 1–0	Denmark wins with 0.60	Croatia versus Nigeria	Croatia 2–0	Croatia wins with 0.59
Egypt versus Uruguay	Uruguay 1–0	Uruguay wins with 0.83	Portugal versus Morocco	Portugal 1–0	Portugal wins with 0.65	Denmark versus Australia	1–1	Denmark wins with 0.72	Argentina versus Iceland	1–1	Draw
Uruguay versus Saudi Arabia	Uruguay 1–0	Uruguay wins with 0.83	Iran versus Spain	Spain 1–0	Spain wins with 0.60	France versus Peru	France 1–0	Draw	Nigeria versus Iceland	Nigeria 2–0	Iceland wins with 0.58
Saudi Arabia versus Egypt	Saudi Arab 2–1	Egypt wins with 0.66	Iran versus Portugal	Portugal 1–1	Portugal wins with 0.64	Australia versus Peru	Peru 2–0	Peru wins with 0.72	Nigeria versus Argentina	Argentina 2–1	Argentina wins with 0.67
Uruguay versus Russia	Uruguay 3–0	Uruguay wins with 0.83	Spain versus Morocco	Spain 2–2	Spain wins with 0.61	Denmark versus France	0–0	Denmark wins with 0.58	Iceland versus Croatia	Croatia 2–1	Croatia wins with 0.74

Table 4 Predicted result versus actual result

E			F			G			H		
Team name	Actual Result	Predicted Result	Team name	Actual Result	Predicted Result	Team name	Actual Result	Predicted Result	Team name	Actual Result	Predicted Result
Costa Rica versus Serbia	Serbia 1–0	Draw	Germany versus Mexico	Mexico 1–0	Germany wins with 0.55	Belgium versus Panama	Belgium 3–0	Belgium wins with 0.72	Colombia versus Japan	Japan 2–0	Colombia wins with 0.67
Brazil versus Switzerland	1–1	Draw	Sweden versus South Korea	Sweden 1–0	Sweden wins with 0.62	Tunisia versus England	England 2–1	England wins with 0.63	Poland versus Senegal	Senegal 2–1	Draw
Brazil versus Costa Rica	Brazil 2–0	Brazil wins with 0.56	Germany versus Sweden	Germany 2–1	Germany wins with 0.60	Belgium versus Tunisia	Belgium 5–2	Belgium wins with 0.55	Japan versus Senegal	Senegal 2–2	Senegal wins with 0.62
Serbia versus Switzerland	Switzerland 2–1	Switzerland wins with 0.59	Mexico versus South Korea	Mexico 2–1	Mexico wins with 0.67	England versus Panama	England 6–1	England wins with 0.79	Poland versus Colombia	Colombia 3–0	Draw
Serbia versus Costa Rica	2–2	Draw	Mexico versus Sweden	Sweden 3–0	Draw	England versus Belgium	Belgium 1–0	Draw	Senegal versus Colombia	Colombia 1–0	Colombia wins with 0.64
Switzerland versus Brazil	Brazil 2–0	Brazil wins with 0.62	Germany versus South Korea	South Korea 2–0	Germany wins with 0.75	Tunisia versus Panama	Tunisia 2–1	Tunisia wins with 0.75	Poland versus Japan	Poland 1–0	Poland wins with 0.71

- This model had to work according to the type of dataset available which could have given better predictions if more accurate data were available.
- In a case study of world cup matches which were concurrently running during this research, this model predicted quite good (63.3%) in group stage matches, but quarter final, semifinal and final prediction were not up to the mark.

Acknowledgements I would also like to show my gratitude to the Authors for sharing their pearls of wisdom with me during the course of this research, and I thank "anonymous" reviewers for their insights.

Compliance with ethical standards

Conflict of interest Md. Ashiqur Rahman declares that he has no conflict of interest.

References

1. Ulmer B, Fernandez M (2014) Predicting soccer match results in the english premier league
2. Maher MJ (1982) Modelling association football scores. Stat Neerl 36(3):109–118
3. Dixon MJ, Coles SG (2002) Modelling association football scores and inefficiencies in the football betting market. J Royal Stat Soc Ser C (Appl Stat) 46(2):265–280
4. Grunz A, Memmert D, Perl J (2012) Tactical pattern recognition in soccer games by means of special self-organizing maps. Hum Move Sci 31(2):334–343
5. Latif A, Rasheed A, Sajid U et al (2019) Content-based image retrieval and feature extraction: a comprehensive review. Math Probl Eng 2019:9658350
6. Ratyal N, Taj IA, Sajid M et al (2019) Deeply learned pose invariant image analysis with applications in 3d face recognition. Math Probl Eng 2019:3547416
7. Kakadiaris IA, Passalis G, Toderici G, Murtuza MN, Lu Y, Karampatziakis N, Theoharis T (2007) Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. IEEE Trans Pattern Anal Mach Intell 29(4):640–649
8. Sajid M, Ali N, Dar SH et al (2018) Data Augmentation-Assisted Makeup-Invariant Face Recognition. Math Probl Eng 2018:2850632
9. Ali N, Zafar B, Iqbal MK, Sajid M, Younis MY, Dar SH et al (2019) Modeling global geometric spatial information for rotation invariant classification of satellite images. PLoS ONE 14(7):e0219833. https://doi.org/10.1371/journal.pone.0219833
10. Sajid M, Ratyal NI, Ali N et al (2019) The impact of asymmetric left and asymmetric right face images on accurate age estimation. Math Probl Eng 2019:8041413
11. Ali N, Bajwa KB, Sablatnig R, Chatzichristofis SA, Iqbal Z, Rashid M, Habib HA (2016) A novel image retrieval based on visual words integration of SIFT and SURF. PloS One 11(6):e0157428
12. Chicco D, Sadowski P, Baldi P (2014) Deep autoencoder neural networks for gene ontology annotation predictions. ACM BCB 2014 - 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, pp 533–540. https://doi.org/10.1145/2649387.2649442.
13. Zell A, Mache N, Huebner R, Mamier G, Vogt M, Schmalzl M, Herrmann KU (1994) Stuttgart neural network simulator. In: Skrzypek J (eds) Neural Network Simulation Environments. The Kluwer International Series in Engineering and Computer Science, vol 254. Springer, Boston, MA
14. Bunker RP, Fayez F (2017) A machine learning framework for sport result prediction. Appl Comput Inf 15(1):27–33
15. Pettersson D, Nyquist R (2017) Football match prediction using deep learning and recurrent neural network applications. Dissertation, Chalmers University of Technology
16. Cireşan D, Meier U, Masci J, Schmidhuber J (2012) Multi-column deep neural network for traffic sign classification. Neural Netw 32:333–338
17. Cireşan D, Meier U, Masci J, Schmidhuber J (2012) Multi-column deep neural networks for image classification. Neural Netw 32:333–338
18. Marblestone AH, Wayne G, Kording KP (2016) Toward an integration of deep learning and neuroscience. Front Comput Neurosci 10:94