

# Topic: Mapping Study of Use of Soft Computing techniques to predict outcomes and team rankings in NFL

M. Razi ul Haq 20100015

M. Adil Inam 20100180

<b>Topic: Mapping Study of Use of Soft Computing techniques to predict outcomes and team rankings in NFL</b>	<b>1</b>
<b>Abstract</b>	<b>1</b>
BACKGROUND	1
OBJECTIVES	2
<b>Introduction</b>	<b>2</b>
Scope and Question	2
Study Selection and Data Extraction	3
<b>Results</b>	<b>4</b>
RQ1: What are the existing approaches to predict the outcomes of sports games?	4
RQ2: Do any of the existing approaches involve Soft Computing techniques? And which ones exactly	4
RQ3: In which sources and in which years were the approaches regarding the prediction of match outcomes published?	4
RQ4: Which sports are particularly used the most in making predictions and what factors allow us to achieve higher prediction accuracy in those sports?	4
<b>Discussion</b>	<b>5</b>
<b>Conclusion</b>	<b>5</b>
<b>References</b>	<b>6</b>

## Abstract

### BACKGROUND

The forecast for winners in sports and the strategies that would make the play successful brings valuable information for both organizers, media and audience. The NFL is the greatest sporting

event in the United States and such information to forecast winning strategies and the number of yards a player would cover could prove to be invaluable. With that in mind, this work presents a study of the main factors influencing match predictability and, a new hybrid approach is proposed to calculate whether or not a touchdown will be successful based on various strategies employed by the team

## OBJECTIVES

We will be looking at all the work done in the field of soft computing regarding predictions, particularly relating to sports results. The aim is to pick out and maintain a record of all the relevant projects that could then further help us at the time of implementation

METHOD: We will be using Google Scholar to limit our search only to papers that have either used Soft Computing techniques to make predictions or papers that have explored prediction of outcomes of sporting events or both of the two things. The abstracts of such papers and their findings and methodology will be analyzed to prevent us from repeating the same work.

## Introduction

### Scope and Question

The project's aim is to accurately forecast the strategies that will make the touchdown successful or not by NFL teams in future by training and using models on the play by play data of past 7 years of data available to us. The model should also incorporate Soft computing techniques wherever possible to increase the accuracy of the predictions. There are two limiting factors:

1. The study should only focus on the last five years of paper.
2. The venues are unlimited but should be restricted to conferences and journals only, published primarily by IEEE, ACM, SPRINGER, ELSEVIER.

No	Research Question	Rationale
RQ1	What are the existing approaches to predict the outcomes of sports games?	The first research question defines the basis of this systematic mapping study.
RQ2	Do any of the existing approaches involve Soft Computing techniques? And which ones exactly	This research question provides an overview of the existing approaches that combine soft computing and conventional techniques used to

		make such classification predictions.
RQ3	In which sources and in which years were the approaches regarding the prediction of match outcomes published?	The third research question indicates whether there are specific publication channels and when effort regarding this research area was made.
RQ4	Which sports are particularly used the most in making predictions and what factors allow us to achieve higher prediction accuracy in those sports?	The fourth research question provides detailed information on the purpose of each approach and what is addressed and improved when applying soft computing approaches.

### Study Selection and Data Extraction

The search for primary studies was based on the following four reference databases: Inspec, Compendex, IEEE Xplore, and ACM Digital Library. The reason for choosing these libraries was that these are some of the most relevant sources in software engineering. The first two databases, Inspec and Compendex, are comprehensive databases containing millions of publications, especially in the engineering and computer science domains.

The aim of the selection process was to identify those articles that are most relevant for the objective of this mapping study. We considered around 10 articles and upon second examination the number was reduced down to 6. These papers were ones that seemed the most relevant one to us when we went through the abstract of each one of these.

A data extraction form was developed to gather all relevant data from the identified articles. Besides general information (e.g., title and author), data with respect to each research question were extracted. This mainly includes a description of the approach (RQ1), the techniques used (RQ2), the publication year and channel (RQ3), evaluation results, goals of the combined approaches (RQ4).

## Results

RQ1: What are the existing approaches to predict the outcomes of sports games?

The articles used a range of methods to make predictions regarding the winning team(classification) and the expected score(regression). Three of the seven articles used Neural Networks including Fuzzy Neuro Networks as well. Others used basic linear regression, logistic regression, decision tree and gradient boosted machine learning.

RQ2: Do any of the existing approaches involve Soft Computing techniques? And which ones exactly

[1], [2], [7] used Soft computing techniques that we may also implement for our predictions. [2] uses an adaptive neural network with back-propagation to make the prediction regarding win-loss outcomes of the team. [7] uses Fuzzy systems and Fuzzy-Neuro systems which are the two most common Soft computing techniques implemented in a range of applications made during the last 5 years. The prediction results of these technologies were almost 78% in the last 10 years of Tennis grand slams.

RQ3: In which sources and in which years were the approaches regarding the prediction of match outcomes published?

The papers are restricted to conferences and journals only, published primarily by IEEE, ACM, SPRINGER, ELSEVIER. Only papers from the past 5 years are used as sources. Fifty per cent of the identified articles present evidence regarding combined approaches. The remaining articles describe just ideas on how a combination could be done or describe combined approaches concretely without giving any evidence about their implementation. To some extent, this can be explained by the fact that many combined approaches have emerged in the past five years and thus, evaluation still has to be done. Furthermore, considering the publication channels, it may be that articles from relevant conferences and journals were not found and thus, certain evaluations were not incorporated into this analysis.

RQ4: Which sports are particularly used the most in making predictions and what factors allow us to achieve higher prediction accuracy in those sports?

The research papers usually cover Basketball. [7] focuses on Tennis and has prediction systems for ATP tennis and World Grand slams. Most of the research is the study of the main factors influencing match predictability and, calculate the chances of victory of each of the competitors before the start of a match. Different techniques are used for example in case of Major league Basketball, statistical methods are used to make predictions based n past trends of teams against other teams. One paper specifically used machine Learning on the dataset for Major Leagues Basketball to make predictions about future win-loss of teams.

## Discussion

The combination of various soft computing techniques such as Fuzzy, Genetic and is one of the most common approaches. Lots of different tools or algorithms are presented for different environments, which shows the heterogeneity of combining static and dynamic analyses. These include lazy learners, artificial intelligence, support vector machines(SVM) and even decision trees. However, it must also be agreed that these are just a handful of research projects on the topic. Usually, the topic of predicting match outcomes in sports is not readily pursued due to the difficulty of making such predictions because of the unpredictability. But the existence of a mechanism that allows these predictions to be made, especially in NBA would prove of immense value particularly if it's accurate as well. Some of the papers employing the fitting accuracy of the improved adaptive BPNN prediction model, multiple linear regression (MLR) model and grey degree prediction model were compared and analyzed. The research results showed that the prediction model built by the improved adaptive BPNN algorithm had a smaller prediction error after rolling prediction. Moreover, this kind of combination is usually conducted in an integrated manner. The articles look at the work done in the field of soft computing regarding predictions, particularly relating to sports results. The current works on this prove that such predictions can be done to high accuracy and that implementing Fuzzy systems give the greatest improvement in the accuracy of such systems. Prediction systems are successfully being used in Basketball and Tennis. One of the articles has also used Strength Equation with weighting factors adjusted by optimization. But mainly all rely on classical techniques of Soft Computing considered relevant for their efficiency and versatility in applications. One interesting technique used for sports prediction was that real-world gambling data was used to train and test different predictive models under varying assumptions. The results show that methods that leverage oddsmaker biases in an exploitative manner perform best under the conditions tested in this paper. These findings suggest that leveraging human and algorithmic decision biases in an exploitative manner may be useful for predicting the outcomes of competitive events

## Conclusion

In this article, the results of a systematic mapping study regarding the implementation of Fuzzy techniques in making predictions about NFL matches and the techniques used were presented. From our point of view, the major result of this systematic mapping study is the identification and classification of existing approaches that use Soft computing techniques. This may support practitioners in selecting worthwhile combinations and serve as a basis for researchers in terms of promising future research directions. We are most interested in a close combination of already existing techniques with Fuzzy, Genetic + Fuzzy and Neuro-fuzzy techniques.

## References

- [1] Ujwal UJ, Antony PJ, Sachin DN (2018) Predictive analysis of sports data using google prediction API. Int J Appl Eng Res 13(5):2814–2816
- [2] Li, H. Analysis on the construction of sports match prediction model using neural network. Soft Comput (2020). <https://doi.org/10.1007/s00500-020-04823-w>
- [3] Brendan Teich and Roman Lutz and Valentin Kassarnig (2016) NFL Play Prediction, arXiv preprint arXiv:1601.00574 (2016)
- [4] Schlicht, Erik J. "Exploiting oddsmaker bias to improve the prediction of NFL outcomes." arXiv preprint arXiv:1710.06551 (2017).
- [5] Bock, J.R. Empirical Prediction of Turnovers in NFL Football. Sports 2017, 5, 1.
- [6] Soto Valero, C. (2016). Predicting Win-Loss outcomes in MLB regular season games – A comparative study using data mining methods, International Journal of Computer Science in Sport, 15(2), 91-112. doi: <https://doi.org/10.1515/ijcss-2016-0007>
- [7] Mateus de Araujo Fernandes. Using Soft Computing Techniques for Prediction of Winners in Tennis Matches. Machine Learning Research.Vol. 2, No. 3, 2017, pp. 86-98. doi: 10.11648/j.mlr.20170203.12