

Challenges/ Factors Affecting Bioinformatic and Machine Learning in Detecting/ Predicting Criminals

Onyekwelu Chidinma Anthonia

Email: onyekweluchidinma6@gmail.com

DOI: 10.56201/ijcsmt.v10.no5.2024.pg114.122

Abstract

In our society today, crimes are committed here and there, within a short distance and interval, without any trace or identification of the criminals and their whereabouts. In most cases these criminals disappear undetected. Criminals parade freely in the society undetected and unknown to the victims of this crime. Some of the victims die or suffer perpetually for the crime these criminals committed. The science of Bioinformatics and Machine Learning is gaining increasingly importance in life, science and technology at large. Bioinformatics and Machine Learning would be a great strategy in eradicating criminals. But it is becoming difficult to detect or predict criminal due to the factors affecting the effectiveness of using bioinformation and machine learning. Crime and Criminal behaviour can be defined by the laws of particular jurisdictions. Numerous problems are affecting the application of bioinformatics and machine learning in detecting and predicting criminals. This research is expected to identify and address the challenges and factors affecting prediction of criminals using Bioinformatics and Machine Learning. The Programming Language suitable for Bioinformatics and Machine Learning for Predicting Criminals will also be discussed in this research. Strategic and careful implementation of the findings will provide the needed solution to these challenges as prescribed identified in the research.

Keywords: Bioinformatics, Crime, Criminal, Detect, Machine Learning, Software, Prediction

Introduction

Bioinformatics is an interdisciplinary field that develops methods and software tool for understanding biological data, in particular when the data sets are large and complex. As an interdisciplinary field of science, bioinformatics combines biology, computer science, information engineering, mathematics and statistics to analyze and interpret the biological data (Gilbert, D. 2004). Bioinformatics is a field of study that uses computation to extract knowledge from biological data. Bioinformatics includes biological studies that use computer programming as part of their methodology, as well as a specific analysis pipeline that are repeatedly used, particularly in the field of genomics (Frank, E. et al. 2004). It includes the collection, storage, retrieval, manipulation and modeling of data for analysis, visualization or prediction through the development of algorithms and software. Bioinformatics deals with computational and mathematical approaches for understanding and processing biological data (Nagesh, S.C. 2019).

It is an interdisciplinary field in which new computational methods are developed to analyze biological data and to make biological discoveries.

Machine learning (ML) is an application that provides a system with the ability to learn and improve automatically from past experiences without being explicitly programmed (Simon, A. et al. 2016). After viewing the data, an exact pattern or information cannot always be determined. ML is applied to interpret the exact pattern and information. ML pushes forward the idea that, by providing a machine with access to the right data, the machine can learn and solve both complex mathematical problems and some specific problems (Dey, A. 2016). Machine learning is used widely in today's world because of increasing computation power and the availability of large datasets on open-source tools (Aashish, U. 2019). ML is categorized into two parts, supervised ML and unsupervised ML. Supervised learning, the machine is trained on the basis of a predefined set of training examples, which facilitates its capability to obtain precise and accurate conclusions when new data are given. Unsupervised learning, the machine is given a set of data, and it must find some common patterns and relationships between the data its own (Musumeci, et al. 2019). In recent years, the size and number of available biological datasets have skyrocketed, enabling bioinformatics researchers to make use of these machine learning algorithms. Machine learning has been applied to six biological domains such as in Genomics, Proteomics, Microarrays, Systems biology, Stroke diagnosis, and Text mining.

Crime is an unlawful act punishable by a state or other authority. The term crime does not, in modern criminal law, have any simple and universally accepted definition, though statutory definitions have been provided for certain purposes. Crime is a category created by law, in other words, something is a crime if declared as such by the relevant and applicable law. A crime or offence or criminal offence is an act harmful not only to some individual but also to a community, society, or the state (Easton, M. 2010). Such acts are forbidden and punishable by law. (Tyagi, D. & Sharma, S. 2018), Crimes have been categories into the following. Forgery, Personation and cheating, Firearms and offensive weapons, Offences against the state/offences against the Crown and Government, or political offences, Harmful or dangerous drugs, Offences against religion and public worship, Offences against public justice, or offences against the administration of public justice, Public order offence, Commerce, financial market and insolvency, Offences against public morals and public policy, Motor vehicle offences, Conspiracy, incitement and attempt to commit crime.

Criminal is a person who has committed a crime. A criminal is someone who breaks the law. Anyone who breaks the law is technically a criminal. Criminal activities are things that are illegal operations. There are different types of criminals which are classified as under such as crimes against a person, crimes against property, inchoate crimes, statutory crimes, financial crimes, habitual criminal, Legalistic criminals, Moralistic criminals, Psychopathic criminals, Institutional criminals or white color criminals, Situational or occasional criminals, Professional criminals, Organized criminals.

Criminal law is the body of law that defines criminal offenses, regulates the apprehension, charging, and trial of suspected persons, and fixes penalties and modes of treatment applicable to convicted offenders. Criminal law is also one of the devices by which organized societies protect

the security of individual interests and ensure the survival of the group (Rummens, A. et al 2017). There are, in addition, the standards of conduct instilled by family, school, and religion, the rules of the office and factory, the regulations of civil life enforced by ordinary police powers, and the sanctions available through tort actions. The distinction between criminal law and tort law is difficult to draw with real precision, but in general one may say that a tort is a private injury whereas a crime is conceived as an offense against the public, although the actual victim may be an individual.

Criminalistics is defined as the application of scientific methods to the recognition, collection, identification, and comparison of physical evidence generated by criminal or illegal civil activity (Wu G. et al. 2003). It also involves the reconstruction of such events by evaluation of the physical evidence and the crime scene. Crimes are committed and no one are held responsible or culprit of the crime. Criminals are all over unknown to the society or to the victims of the crime. Some of the victims of these crimes die or suffer perpetually for the crimes these criminals committed. Is not fair that these criminals should be moving freely without being caught, identified or detected and they are living, moving freely in the society (Panchiwala& Shah M. 2020). It is even funny at times to find out that these criminals are always at the consoling point of their victims and equally suggests way of fetching the culprits. God have mercy on us, image victims dinning unknowingly with their traitors.

The crime committed in our society nowadays are quite sympathetic and unimaginable. Victims at times report the crime to the police stations. The case may not be handled properly. Still the victims may end up been intimidated by these same police that supposed to defend and protect life and property (Wu G. et al. 2003). The victim may even be converted to the criminal of the crime committed against them. The case may equally end up not been attended to or be in favour of the criminal of the crime.

These criminals can be easily detected when the challenges affecting bioinformatics and machine learning in detecting and predicting criminals are overcome (Walczak, S. 2021). Once these challenges are tackled, criminal will be easily detected, predicted and crime will reduce in the society. Because once the criminals find out that their colleagues are easily identify, they will disease from crimes and environment. Then our society will become crime free and peaceful. Criminals will equally evacuate our society.

Problem Statement/Justification

Our society is saturated with all sorts of crime and criminals. The rate, time and how crimes are committed are so sympathetic. Today, crimes are committed here and there, within a short distance and interval, without any trace or identification of the criminals and their whereabouts. Some of the identified crimes are broad-day robbery, snatching of cars, snatching of phones, hand bags and bags etc. Breaking into houses, shops to steal goods and properties are daily routine, raping and kidnapping just to name but a few. The most annoying aspect of it all, is that these criminals go scort free without been detected or apprehended. In most cases these criminals disappear undetected. These criminals parade freely in the society undetected and unknown to the victims of this crime. These criminals enjoy and lavish the ill-gotten fortune in the same society and even

claim to be richest in the society. Some of the victims die or suffer perpetually for the crime these criminals committed. The time interval of reporting and searching for the criminal is quiet disgusting, with respect to the law enforcement agencies. Tracking of the criminals and their movement is zero because of bad road, network and technology involved in our society. Engaging the police and law court is another questionable situation that would not be sure of the outcome, even being the victim of the circumstances.

Objectives of the Study

The objectives of this study will gear towards the following:

- I. To identify the challenging and factors effecting bioinformatics and machine learning in predicting and detecting criminals.
- II. To suggest solution to the identified challenges affecting bioinformation and machine learning in predicting and detecting criminals
- III. To identify, develop and suggested Programming Language suitable for developing bioinformatics and machine Learning for predicting criminals.

Literature Review

Behavior analysis and crime prediction using big data and machine learning (Jha P. et al. 2019), is a big data and ML technique for behavior analysis and crime prediction. The authors discuss the tracking of information using big data, different data collection approaches, and the last phase of crime prediction using ML techniques based on data collection and analysis. They presented a predictive analysis through ML using RapidMiner by processing historical crime patterns. it was conducted in four phases: data collection phase, data preparation p, data analysis phase, and data visualization phase. Big data was concluded as the suitable framework for analyzing crime data. The main reasons are as follows: it can analyze extremely large datasets, and generate reliable results, whereas the ML based naive Bayes algorithm can achieve better predictions using the available datasets and provide a high throughput and fault tolerance.

In research, how to test bioinformatics software? (Amir, H.K & et al.), the authors discussed Bioinformatics as an application of computational, mathematical and statistical techniques to solve problems in biology and medicine. Bioinformatics programs developed for computational simulation and large-scale data analysis are widely used in almost all areas of biophysics. The appropriate choice of algorithms and correct implementation of these algorithms are critical for obtaining reliable computational results. Nonetheless, it is often very difficult to systematically test these programs as it is often hard to verify the correctness of the output, and to effectively generate failure-revealing test cases. Software testing is an important process of verification and validation of scientific software, but very few studies have directly dealt with the issues of bioinformatics software testing. This work reviewed important concepts and state-of-the-art methods in the field of software testing. it also discusses recent reports on adapting and

implementing software testing methodologies in the bioinformatics field, with specific examples drawn from systems biology and genomic medicine.

In research, Design and analysis of machine learning algorithms for the reduction of crime rates in India (Bandeekar SR, & Vijayalakshmi C. 2020). They focused on the analysis and design of ML algorithms to reduce crime rates in India. ML techniques were applied to a large set of data to determine the pattern relations between them. The research was mainly based on providing a prediction of crime that might occur based on the occurrence of previous crime locations. Techniques such as Bayesian neural networks, the Levenberg Marquardt algorithm, and a scaled algorithm were used to analyze and interpret the data, among which the scaled algorithm gave the best result in comparison with the other two techniques. Based on the correlation, analysis of variance, proved that with the help of the scaled algorithm, the crime rate can be reduced by 78%, implying an accuracy of 0.78 were statistically analysed.

In this project, Machine Learning Approach to Crime Prediction and Identification of Hotspots, (Amshumann, S. 2021), delves into the crime dataset for the city of Denver and attempts to identify hotspots and types of crime occurring in the city. The FP-growth algorithm helps to identify the location and time at which crime would be likeliest to occur in the city, which can help in police monitoring and faster response teams in the area. Using the classification algorithm, we can try to guess the type of crime occurring which allows the police and the government to take action and better deal with crime. The neural network model worked best for this purpose but if the objective is to have reasonable probability over the different categories of crime, random forest is more effective.

As the accuracy metric indicates, it's quite clear that any of the classification models cannot actually be applied in a real-world scenario. There is clear scope for improvement in utilizing the data to make improved predictions.

In a research, Real time hand gesture recognition system for dynamic applications (Rautaray, SS. 2013). Because many at times, during criminal activities, convicts use hand gestures to signal messages to each other. Hand gesture recognition was conducted using computer vision models. It can be achieved by capturing images, and then try detecting a hand in the background. Then apply either computer aided manufacturing or different procedure in which they first convert a picture into gray scale, after which they set the image return on investment, and then find and extract the biggest contour. And then determine the convex hull of the contour to try and find an orientation around the bounded rectangle. The gesture and were converted it into a meaningful command and were interpreted.

In a paper, Detecting Fault Modules Using Bioinformatics Techniques, many software reliability studies attempt to develop a model for predicting the faults of a software module because the application of good prediction models provides important information on significant metrics that should be observed in the early stages of implementation during software development (Nagesh S. C. 2019). They proposed a new method which was inspired by a multi-agent-based system that was initially used for classification and attribute selection in microarray analysis. Best classifying gene subset selection is a common problem in the field of bioinformatics. If the software metrics

measurement values of a software module were regard, as a genome of that module, and the real-world dynamic characteristic of that module as its phenotype i.e. failures as disease symptoms, we can borrow the established bioinformatics methods in the manner first to predict the module behavior and second to data mine the relations between metrics and failures

An efficient automated disease diagnosis model was designed using the machine learning models (Kumar, et al). Is a research that discuss on the three critical diseases such as coronavirus, heart disease, and diabetes were selected. In the proposed model, the data are entered into an android app, the analysis is then performed in a real-time database using a pretrained machine learning model which was trained on the same dataset and deployed in firebase, and finally, the disease detection result is shown in the android app. Logistic regression is used to carry out computation for prediction. Early detection can help in identifying the risk of coronavirus, heart disease, and diabetes. Comparative analysis indicates that the proposed model can help doctors to give timely medications for treatment. The machine learning models are applied to the coronavirus, heart disease, and diabetes dataset to predict the risk of these diseases in an individual. An end-to-end process is used where people must enter their details in the mobile application and submit the data. The real-time processing takes place, and the risk is predicted within a few seconds. The mobile application that is used as a real-time database on the cloud is the firebase database. The trained parameters of the model are stored in the database, and prediction is done in real-time. The main contributions are: (i) An efficient automated disease diagnosis model is designed using the machine learning models. (ii) Three critical diseases are selected such as coronavirus, heart disease, and diabetes. (iii) In the proposed model, the data are entered into an android app, the analysis is then performed in a real-time database using a pretrained machine learning model which was trained on the same dataset and deployed in firebase, and finally, the disease detection result is shown in the android app. (iv) Logistic regression is used to carry out computation for prediction.

In research, a video analysis framework for soft biometry security surveillance, (Wang YF, et al. 2005). It is designed for identifying people for surveillance with the help of a soft biometry. It identifies a person's height, built, skin tone, shirt and trouser color, motion pattern, and trajectory history to identify and track passengers, which can equally help in predicting crime activities. Some absurd human error incidents that have resulted in the perpetrators getting away were also discussed. Experiments were conducted, the results of which were quite impressive. The cameras will be able to show airport guards and passengers at their dust post.

In research, a graphical user interface-based prediction of crime rates using a ML approach is presented (Prithi, et al. 2020). The main focus of this study was to investigate machine-learning-based techniques with the best accuracy in predicting crime rates and explore its applicability with particular importance to the dataset. Supervised ML techniques were used to analyze the dataset to carry out data validation, data cleaning, and data visualization on the given dataset. The results of the different supervised ML algorithms were compared to predict the results. The system consists of data collection, data preprocessing, and construction of a predictive model, dataset training, dataset testing, and a comparison of algorithms. The aim of this study is to prove the effectiveness and accuracy of a ML algorithm for predicting violent crimes.

Methodology

This research tends to discuss: Meaning of Bioinformatics. Meaning of Machine Learning, Crime and Criminal. It will explore primary and secondary data to effectively identify the challenges/Factors affecting Bioinformatic and Machine Learning in Detecting/Predicting Criminals. The Programming Language suitable for Bioinformatics and Machine Learning for Predicting Criminals. The suggested solution to the challenges. The research will be focused on the development of software that will be able to track an impending crime.

The materials needed for this research will be secondary data such as published textbook, peer-review journals and other electronics materials.

Results

It is expected to identify in-depth knowledge of bioinformatics and machine learning and their applications in determining criminals in their different capacities. Both applications (bioinformatics and machine learning) are expected to have their individual limitations and when both applications are combined, they would overcome their individual limitations. The value of machine learning and bioinformatics are rooted in its ability to create accurate models to guide future actions and to discover patterns that have never seen before in predicting and detecting criminals. The process of uncovering criminal activity or verifying reported crime and acquiring evidence in order to identify and prosecute the perpetrators would be easily achieved with application of bioinformatics and machine learning. Machine Learning brings endless possibilities to Bioinformatics. This research discovered that these applications have numerous and great challenges or factors affecting predicting or detecting criminals such as Time Factor, track and report, movement, gathering of evidence, search and arrest of suspects, bias trial procedural system, interrogation and confession, identification of suspect, health and technological issues.

References

- Aashish Upadhyay. (2019). Crime Detection Using Machine Learning. Software Engineer at S&P Global published 25, July 2019
- Amir Hossein Kamali, Eleni Giannoulatou, Tsong Yueh Chen, Michael A Charleston , Alistair L McEwan, Joshua W K Ho. How to test bioinformatics software?
- Amshumann Singh, (2021). Machine Learning Approach to Crime Prediction and Identification of Hotspots. Published Aug 30, 2021.
- Bandekar SR, Vijayalakshmi C. (2020). Design and analysis of machine learning algorithms for the reduction of crime rates in India. *Procedia Comput Sci.* 2020; 172:122–127. doi: 10.1016/j.procs.2020.05.018. [CrossRef] [Google Scholar]
- Bharati A, Sarvanaguru RAK. (2018). Crime prediction and analysis using machine learning. *Int Res J Eng Technol.* 5(9):1037–1042. [Google Scholar]
- Dey, A (2016). Machine learning algorithms: a review. *Int J Comput Sci Inf Technol* 7(3):1174–1179.

- Easton, Mark (17 June 2010). What is crime? BBC News. Archived from the original on 27 February, 2013. Retrieved 10 June 2013.
- Frank E, Hall M, Trigg L, Holmes G, Witten IH. (2004). Data mining in bioinformatics using Weka. *Bioinformatics*. 20(15):2479–2481. doi: 10.1093/bioinformatics/bth261. [PubMed] [CrossRef] [Google Scholar]
- Gilbert D (2004). Bioinformatics software resources. *Briefing in Bioinformatics*. 5 (3): 300–304. doi:10.1093/bib/5.3.300. PMID 15383216.
- Jha P, Jha R, Sharma A. (2019). Behavior analysis and crime prediction using big data and machine learning. *Int J Recent Technol Eng*. 2019; 8(1):461–468. [Google Scholar]
- McClendon L, Meghanathan N. (2015). Using machine learning algorithms to analyze crime data. *Mach Lear Appl Int J*. 2015; 2(1):1–12. doi: 10.18642/ijamml_7100121446. [CrossRef] [Google Scholar]
- Musumeci F, Rottondi C, Nag A, Macaluso I, Zibar D, Ruffini M. (2019). An overview on application of machine learning techniques in optical networks. *IEEE CommunSurv Tutorials* 21(2):1381–1408. <https://doi.org/10.1109/COMST.2018.2880039>.
- Nagesh S. C. (2019). Explore the world of Bioinformatics with Machine Learning. *Bioinformatics, Machine Learning, Python*. KDnuggets on September 17, 2019
- Naresh Kumar, Nripendra Narayan Das, Deepali Gupta, Kamali Gupta, and Jatin Bindra (2021), Efficient Automated Disease Diagnosis Using Machine Learning Models. Published 06 May 2021
- Panchiwala S, Shah M. (2020). A comprehensive study on critical security issues and challenges of the IoT world. *J Data Inf Manag*. 2(7):257–278. doi: 10.1007/s42488-020-00030-2. [CrossRef] [Google Scholar]
- Prithi S, Aravindan S, Anusuya E, Kumar AM. (2020). GUI based prediction of crime rate using machine learning approach. *Int J Comput Sci Mob Comput*. 9(3):221–229. [Google Scholar]
- Tyagi D, Sharma S. (2018). An approach to crime data analysis: a systematic review. *Int J Eng Technol Manag Res*. 5(2):67–74. [Google Scholar]
- Rautaray SS. (2012). Real time hand gesture recognition system for dynamic applications. *Int J Ubi Comp*. 2012;3(1):21–31. doi: 10.5121/iju.2012.3103. [CrossRef] [Google Scholar]
- Simon A, Deo MS, Venkatesan S, Babu DR (2016) An overview of machine learning and its applications. *Int J Electr Sci Eng* 1(1):22–24.
- Walczak, S. (2021). Predicting Crime and Other Uses of Neural Networks in Police Decision Making. *Front. Psychol*. 12:587943. doi: 10.3389/fpsyg.2021.587943

- Wang YF, Chang EY, Cheng KP. (2005). A video analysis framework for soft biometry security surveillance. Paper presented at the 3rd ACM international workshop on video surveillance & sensor networks. Hilton: ACM. [Google Scholar]
- Wu G, Wu Y, Jiao L, Wang YF, Chang EY. (2003). Multi-camera spatio-temporal fusion and biased sequence-data learning for security surveillance. Paper presented at the 11th ACM international conference on multimedia. Berkeley: ACM. [Google Scholar]