# Short-term Traffic Flow Prediction of Urban Roads Based on Random Forest

Luofeng Jiang
School of Software Engineering
Chengdu University of Information Technology
Chengdu,China

**Abstract**: In the intelligent traffic system, accurately grasp the intrinsic changes in the traffic flow law, and timely scientific prediction of the future traffic flow for a number of moments, will have a very important significance to traffic guidance work, traffic management work, traffic planning work, etc. In this paper, we propose a short-time traffic flow prediction model based on random forest regression(RFR) algorithm, and improve the performance of the model by adjusting the hyperparameters of the model. The performance of the random forest model was also compared with the support vector regression (SVR) model and the decision tree regression(DTR) model, and the RFR model ultimately yielded the best predictions.

**Keywords**: Prediction; Random forest; Regression; Traffic flows;Model

## 1 INTRODUCTION

With the development of society and economic progress, a series of traffic problems have become increasingly serious problems that many cities have to face, and the application of intelligent transportation systems to ease traffic congestion has become possible. Among them, short-term traffic flow prediction, which is one of the important foundations of intelligent transportation system, is the key to real-time, accurate and rapid traffic management, inducement and control [1]. Therefore, the study of short term traffic flow prediction on roads is of great significance to improve the level of urban traffic management. At present, there are many models and methods that can be used for traffic flow prediction, commonly used methods such as historical mean model [2], time series model [3], Kalman filter model [4] and other prediction methods and models based on traditional mathematics, as well as prediction methods based on neural networks [5], wavelet theory [6], non-parametric regression [7], support vector machine [8] and other non-linear theories.

This paper presents a model based on random forest regression to predict short term traffic flow on roads. Because random forests are characterized by nonlinear mapping ability, self-learning and adaptive ability, generalization ability and fault tolerance, the predictive accuracy of the model is guaranteed while reducing model complexity, overfitting and computational volume.

## 2 RANDOM FOREST

Random forest has both regression and categorization, and when the study variable is a continuous variable, random forest regression is used for analysis, and when the study variable is a categorical variable, random forest classification is used for analysis. The traffic flow data samples in this paper are continuous variables and fall under the scope of regression methods.

Random Forest is an integrated learning method based on Bagging. The algorithm refers to the analytical prediction of a sample by constructing a combined model through multiple decision trees, each decision tree model will have a predictive value, and the predictive value of each tree will be aggregated, which will ultimately enhance the predictive effect of the model. The regression prediction will average the predicted value of each tree to get the final predicted value.

The random forest regression algorithm process is as follows.

(1) Random generation of sample subsets using

Bagging ideas.

(2) K attributes are randomly selected from a large number of attributes, node splitting is performed, and a single regression decision sub-tree is constructed.

(3) Repeat steps 1 and 2 to construct N regression decision sub-trees to form a forest.

(4) The predicted values of the N decision sub-trees are averaged as the final predicted result.

In the above algorithmic process, no pruning is required during the splitting of each decision tree.

## 3 RESULTS AND ANALYSIS

This section begins with an introduction to data sources and data collection. The process of constructing the prediction model is then described, the parameter adjustments in the construction process are described in relation to it, and the performance of the prediction model is evaluated by the prediction results.

### 3.1 Experimental Data

In this paper, the subject of the study is an intersection in Mianyang City, the time period is November 4, 2019 to November 29, 2019, remove the weekend, a total of 20 days of traffic flow data, every 5 minutes is a time period, the whole data set is processed statistics to get 5760 sets of data, split it into training set and test set, training set 4608 sets, test set 1152 sets.

### 3.2 Model Performance Metrics

A common regression model accuracy is used as a metric for parameter tuning and performance analysis of random forest regression models. The meaning absolute error(MAE), R-squared($R^2$) and root mean square error(RMSE) are included. the smaller the MAE and RMSE, the larger the $R^2$, the higher the model accuracy.

$$MAE = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}|\hat{y}_i - y_i| \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)}{\sum_{i=1}^{N}(y_i - \overline{y_c})} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \quad (3)$$

In equation (1), equation (2) and equation (3), $y_i$ is the true value; $\hat{y}_i$ is the predicted value; $\overline{y_c}$ is the mean of the true value; N is the sample size.

### 3.3 Model Parameters

After initializing the random forest regression model, a grid search was applied for parameter tuning and 10-fold cross-validation (Python based scikit-learn package) to get the parameters min_samples_split is 12, min_samples_leaf is 1, max_features is 29 and max_depth is None.

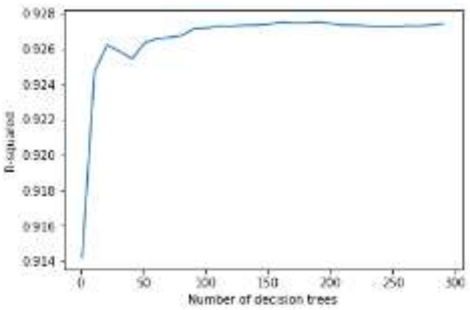


Figure 1. Learning curve a

Figure 1 shows the change in performance when the number of decision trees in the random forest regression model is increased by 10 decision trees at a time from 1-300, and it can be seen that the accuracy is highest when the number of decision trees is around 190. Figure 2 represents the change in performance when the number of decision trees in the random forest regression model is increased by one decision tree at a time over the range 182-201, it can be seen that the model has the highest accuracy when the number of decision trees is set at 186, the parameter n_estimators is 186.

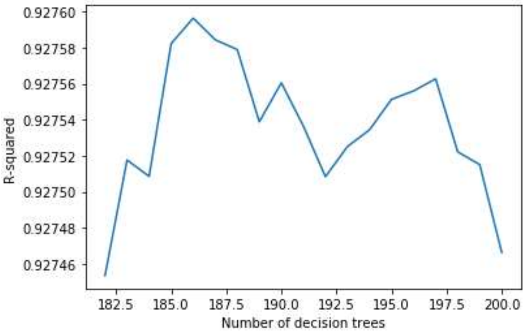


Figure 2. Learning curve b

### 3.4 Comparative Model Analysis

Calculate the MAE, RMSE and $R^2$ of the model according to the training results of the model on the experiment, and compare the experimental model with the support vector regression prediction model and the decision tree prediction model. The results are shown in Table 1.

**Table 1. Model comparison**

| | $R^2$ | MAE | RMSE |
|---|---|---|---|
| RFR | 0.927 | 17.213 | 26.436 |
| SVR | 0.926 | 18.437 | 26.669 |
| DTR | 0.888 | 20.528 | 32.798 |

By comparing the $R^2$, MAE and RMSE of three models of decision tree regression, support vector regression and random forest, the prediction effect of random forest regression is finally shown to be better.

## 4 CONCLUSION

Random forest, as an efficient machine learning algorithm, has been widely used in many fields. However, there are few applied researches in traffic flow prediction. In this paper, through the research and implementation of the road short-term traffic flow prediction model based on random forest regression algorithm, it is found that the model has high prediction accuracy.

## 5 REFERENCES

[1] VLAHOGIANNI E I，GOLIAS J C，KA R LAFTIS M G. Short-term Traffic Forecasting: Overview of Objectives and Methods ［J］. Transport Reviews，2004，24 ( 5 ) :533 — 557

[2] Okutani I, Stephanedes Y J. Dynamic prediction of traffic volume through Kalman filtering theory [J]. Transportation Research Part B: Methodological,1984,18(1): 1-11.

[3] KARLAFTIS M G， VLAHOGIANNI E I. Memory Properties and Fractional Integration in Transportation Time Series ［ J ］. Transportation R esearch Part C:Emerging Technologies，2009，17 ( 4) : 444 — 453.

[4] GUO J H，WILLIAMS B M. Real Time Short Term Traffic Speed Level Forecasting and Uncertainty Quantification Using Layered Kalman Filters ［J］. Transportation Research Record，2010，2175: 28 — 37

[5] Ledoux C. An urban traffic flow model integrating neural networks [J]. Transportation Research Part C: Emerging Technologies, 1997, 5(5): 287-300.

[6] YANG Chun-xia， FU Yi-qin， BAO Tie-nan. Short-term Traffic Flow Prediction Based on Similarity ［J］. Journal of Highway and Transportation R esearch and Development，20015，32 ( 10) : 124 — 128.

[7] ZHANG Tao，CHEN Xian，XIE Mei-ping，el al. K-NN Based on Nonparametric Regression Method for Short Term Traffic Flow Forecasting ［J］. System Engineering-Theory & Practice，2010，30 ( 2) : 376 — 384.

[8] WU Q. A Hybrid-forecasting Model Based on Gaussian Support Vector Machine And Chaotic Particle Swarm Optimization ［J］. Expert Systems with Applications，2010，37 ( 3) : 2388 — 2394.

International Journal of Computer Applications Technology and Research
Volume 9–Issue 05, 172-176, 2020, ISSN:-2319–8656

# Automated Crime Patterns Analysis Framework for Predictive Policing using Data Mining Techniques

Duncan Nyale
Directorate of Computing and e-Learning
The Cooperative University of Kenya
Nairobi, Kenya

Michael M Kangethe
e-Kraal
Cyber Security Innovation Hub
Nairobi, Kenya

**Abstract**: The aim of this research is to study and develop an automatable technological framework that can be used to identify contributing attributes, patterns and trends from reported cases using data mining techniques. A combination of classification and association rules based data mining approach has been proposed for this study due to its effectiveness in bringing out patterns and trends that are interlinked, related and near each other**.**

**Keywords**: ANN: - Artificial Neural Network, KNN: - K Nearest Neighbour, MAP: - Maximum A Posteriori, GUI: - Graphical User Interface

## 1. INTRODUCTION

Through the digitization of reported cases by several law enforcement and public oversight agencies, the need for faster and reliable methods of sifting through massive data and cases to identify attributes and patterns that could lead to a future occurrence arises. Currently when policing and conducting security operations in response to reported crimes most agencies either use the previous reports and will have to sift through a lot of records to find patterns or conduct blanket policing which both are inefficient and laborious. This research details a general framework developed from the use of a combination of several data mining techniques to map occurrences to their dominant attributes and combinations.

In this age of vast data generation, it is imperative that we should find new and novel ways of effectively and quickly analyzing data and give appropriate feedback for decision making in any sphere. The increased computing power coupled with artificial intelligence and machine learning can be used for data mining, or knowledge discovery in databases to bring out previously unknown and potentially useful information from data. Intelligent data analysis is to extract useful knowledge, a process which demands a combination of several things including extraction, analysis, conversion, classification, organization and reasoning. This is precisely what this research has managed to do by creating an intelligent analysis framework that can be universally applied to any data by combining two data mining techniques to leverage on both to create a reliable model applicable to law enforcement through intelligence based policing.

## 2. BACKGROUND CONCEPT

***Data mining*** is a relatively new data analysis technique that has the ability to discover patterns stored within historical data and is now considered a catalyst for enhancing business processes by avoiding failure patterns and exploiting success patterns. Several data mining techniques have been developed over the last decade. Generally, the data mining techniques can be categorized in four categories, depending on their functionality: classification, clustering, numeric prediction, and association rules. The main difference between the different techniques is in the way they extract information (algorithms and methods used) and how results (knowledge discovery/rules) are expressed. ( Khaled Nassar, March 2007)

***Instance Based Learning***: This is the approximation of the target function from the training examples, as the approximation process is repeated with each and every query. Each time a new instance is encountered, its relationship to the previously stored examples is examined to assign a target function value for the new instance. There are several algorithms which include the Locally Weighted Regression, Case Based Reasoning, and the one to observe, the K- Nearest Neighbor and the Radial Basis Functions.

***Classification*****:** problems are essentially predictive models used to analyze an existing database to determine categorical divisions or patterns in the data. Classification problems are focused on identifying the characteristics indicating the group or class to which each record in the database belongs. On the other hand, when there is no pre-identified class or group, the clustering technique is used to group items that seem to fall naturally together. Several algorithms are inherently designed and suited for this purpose which include the KNN, ANN, Radial Basis Functions

The data mining technique used in this research will be a combination of instance based association and classification machine learning algorithms. In association learning, the goal is to discover any interesting patterns in the data by discovering association rules. Association rules differ from classification rules in two ways: they can predict any attribute (not just the group or class), and they can predict more than one attribute's value at a time. A typical association rule is represented in the following way:

Cause_1, Cause_2 => Result (or consequence)

That is, if Cause_1 and Cause_2 hold then Result (the association rule) applies, for n% of cases with x% confidence.

Each rule extracted is usually provided with a confidence level and a support. The confidence is the statistical value presenting the probability of a certain rule and the support is the number of cases/projects in which the rule is found. A pattern is defined as several identical or similar rules indicating a trend. Most of the data mining techniques use statistical tests when constructing rules or patterns and also for correcting models that depend too strongly on particular records in producing the rules and patterns *(Feldens 2002).* Since the goal when analyzing the dataset collected here was to detect any potentially useful patterns within the target industry based on reported incidences and registered complaints, association learning was the data mining technique selected to analyze the dataset collected in this research.

*(Bruno Agard, Catherine Morency and Martin Trépanier 2007)* Conducted a research on data mining methords for the transport industry user behavior using Smart Card Data with desirable results. The limitations to their research based on the proposed research were that theirs focused on commuters behavioral patterns for economic and financed based planning of their transport system. Their observations during their research showed that the public transport users of this study can rapidly be devided into four major behavioural groups, whatever type of tickets used.

*(Vikas.Grover et al 2009*) Examined the current techniques that are used to predict crime and criminality. They were able to narrow down their research of possible techniques to three main categories:

- ***Statistical Methods***, these mainly relate to the journey to crime, age of offending and offending behavior.
- ***Techniques using Geographical Information Systems*** that identify crime hot spots, repeat victimization, crime attractors.
- ***Crime Generators;*** a miscellaneous group which includes machine learning techniques to identify patterns in criminal behavior and studies involving reoffending. Although their research provided a great insight on methords of crime patterns analysis. Their approach was not focused on any particular industry and thus followed no formal government policy.

## 3. METHODOLOGY

The design of this prototype to reflect the system framework will be done in layers and components which will be designed individually and developed independently to reflect the independent framework sub processes. Each subcomponent will take a collection of different discreet valued, real valued, inputs and produce real valued outputs that will at times be the input values of other or the same components based on the instance of computation.

The main framework will be based on a hybrid model derived partially from the combination of the two case based reasoning techniques which is the Bayesian Belief Networks from the Bayesian Based learning class of algorithims, the main concept of the Bayesian Based Learning which in principle is as below.

Features of Bayesian learning methods:

- Each observed training example can incrementally decrease or increase the estimated probability that a hypothesis is correct, unlike algorithms which completely eliminate a hypothesis if it is inconsistent with any single example
- Prior knowledge can be combined with observed data to determine the final probability of a hypothesis. Prior probability is provided by asserting a prior probability for each candidate hypothesis and also a probability distribution over observed data for each possible hypothesis
- Bayesian learning can accommodate hypotheses which make probabilistic predictions e.g. this patient has a 93 % chance of recovery
- New instances can be classified by combining the predictions of multiple hypotheses, weighted by their probabilities. The main principle borrowed from this methord is the Maximum A Posteriori (MAP) Hypothesis and Maximum Likelihood in which the main goal of the methord is To find the most probable hypothesis ***h*** from a set of candidate hypotheses ***H*** given the observed data ***D***.

***MAP Hypothesis,*** $h_{MAP} = \mathrm{argmax}_{h\in H}\, P(h|D)$

$$= \mathrm{argmax}_{h\in H}\, P(D|h)P(h)/P(D)$$

$$= \mathrm{argmax}_{h\in H}\, P(D|h)P(h)$$

Where If every hypothesis in *H* is equally probable a priori, we only need to consider the *likelihood of the data D* given *h*, $P(D|h)$. Then, $h_{MAP}$ becomes the ***Maximum Likelihood***, $h_{ML} = \mathrm{argmax}_{h\in H}\, P(D|h)$

in essence the formular is explained as below:

- To determine the most probable hypothesis, given the data *D* plus any initial knowledge about the prior probabilities of the various hypotheses in *H.*
- ***Prior probability of h, P(h):*** it reflects any background knowledge we have about the chance that *h* is a correct hypothesis (before having observed the data – independent of D). If ***P(h)*** is not known, we can assign same probability to each hypothesis
- ***Probability of D, P(D):*** it reflects the probability that training data *D* will be observed given no knowledge about which hypothesis *h* holds.
- ***Conditional Probability of observing D, P(D|h):*** it denotes the probability of observing data *D* given some world in which hypothesis *h* holds.

And Concept Learning which is Inferring a boolean-valued function from training examples of its input and output.

Concept learning can be formulated as a problem of searching through a predifined search space of potential hypothesis for the hypothesis that best fits the training examples. The hypothesis might be represented in the form below where:

- Hypothesis h is a conjunction of constraints on attributes
- Each constraint can be:
  - ■ A specific value : e.g. *Gender=Male*
  - ■ Range Value e.g. *Age=19-25*
  - ■ A don't care value : Marital Status=?
  - ■ No value allowed (null hypothesis): e.g. *Height*

Example: hypothesis *h*

| Time | Gender | Marital Status | Location | Age | Crime |
|---|---|---|---|---|---|
| 1800 - 1900Hrs | M | Single | Nairobi | 19-25 | Armed Robbery |
| 0.9 | 0.99 | 0.3 | 0.6 | 0.8 | |

The main approach to this research will be in two main stages*;*

i. Theoretical model framework design and

ii. System prototype development and model testing

## 3.1 Theoretical Model Framework Design

This will be done in three sub-stages as structured below.

**Framework research and design**

This is the first stage of the overall research as it will involve the research of existing systems. Identification of successful designs in related areas of application. Identification of the relevant frameworks approach development of discreet overall steps and stages of the proposed framework.

The outcome of this stage will be a high level description of the system framework that will be used to solve the said problem. It will be a diagrammatic representation of the whole system framework as a flowchart with generalized descriptions of the individual components that make up the final system.

**Input data structure and format design**

This is the second sub-stage as it will involve the identification of all the possible outcomes (which in this case will be cases or reported incidences). The identification of all attributes that affect the outcomes. And also the structure of the attributes and their significance in how and by what magnitude they affect the effective outcome. From there the identified attributes will have to be converted into discreet or semi-discreet value or variable based data inputs so that the proposed model will be able to mathematically compute the outcomes based on the associated attributes.

The outcome of this stage will be the creation of a discreet and continuous value based attributes input data structure that will eventually dictate the final systems database design.

This stage will overlap with the third sub-stage which is the Mathematical model design as the model will dictate the data and type of data needed as model inputs.

**Mathematical model design**

This will be the final stage of the theoretical system framework research process. This stage will involve the actual mathematical computation model formulation. It will involve “mapping” the attributes to their occurrences and identification of their associations to the overall result outcomes. Each attribute will be analyzed separately and together with all others to identify the magnitude effect on the overall outcome they have. Once the attributes, their relationships and effects to the final outcome have been discretely identified, they will be combined with respect to their magnitude and effect to the overall outcome to provide the actual mathematical formula that will be mapped on to the framework to provide the complete theoretical blueprint of the system that can be implemented by a variety of different programming technologies and approaches that observe the algorithmic process. This process will involve several mathematical processes. The processes will focus on two main things.

1. Attribute incident association calculation.

Each associated attribute impacts the overall outcome differently as some have a greater impact on the outcome than others. This will be achieved by computing the magnitude effect of the attribute individually and by combination with other attributes that result to the same outcome.

The general mathematical rule to be observed will be as below:

$$w_{c_a} = \sum_{1}^{n} a_{c_i}$$

Equation 1: Proposed Attribute incident equation

Where:

- $w_{c_a}$ Is the overall weight of the attribute in terms of how it affects the outcome.
- $n$ Is the number of possible combinations of the attributes.
- $a_{c_i}$ Is the combination instance of the attribute combination.

This initial model is subject to alteration and improvement during the research process. The result of this model will be the mathematical determination of the impact of specific attributes or conditions to an outcome.

2. Attribute Effect Calculation

This model will also be used to identify the most dominant attribute that leads to an occurrence or incident. The reverse association rule will be as below:

$$a_{Dominant} = a_{i_{MAX}}$$

Equation 2: Attribute Effect Magnitude Equation

This means the most dominant attribute that contributes to an occurrence is that attribute that has the highest number of correlation values in the said instance of computation.

Where:

- $a_{Dominant}$ is the dominant attribute
- $a_{i_{MAX}}$ is the particular attribute instance *i* of all the available attributes

This first main stage will involve the most research effort and time of implementation as the whole concept and purpose of the research is based on this stage.

## 3.2 Model Testing and Evaluation

It will involve the actual development of a “proof of concept” system prototype that will not only be independently testable but will demonstrate the practicability of the whole research as an artificially intelligent computer system. This stage will involve several structured independent and dependent sub-stages that will be followed to come up with an actual working model of the system. The stages involved will be as follows.

**Database and input and output data structure design**

This will be the actual database design process of the overall system. The proposed database technology to be used will be the MySQL database Technology. Reasons for following this approach are due to the fact that MySQL has the following but not limited to, advantages.

- It is FREE: one does not need a license to use this technology as no costs are needed to implement this technology.

- It is SCALABLE: the database can grow and accommodate large volumes of data at an exponentially increasing rate.
- It is FLEXIBLE: the ease of implementation of this database to different data structure rules and methods is almost seamless as the technology is not rigid by nature.
- Easy to use: the amount of technical skill needed to implement this database technology is minimal as there is enough literature on nearly every imaginable scenario.

**Prototype Interface design**

This stage will involve the actual system GUI design. It be done in two stages

1. Input control design

This sub-stage will involve the actual design of the main system user interface and the required controls needed to for the easy interaction and manipulation by the user in computing the output. The main focus of this stage is to design the controls and selection rules needed to compute the final outcome based on the selection criteria. This is where the user will interact with the system in performing the analysis.

2. Output results design

This sub-stage will involve_the designing of the results presentation interface. Factors to be considered in this stage are the ease of interpretation of the results in such a way a layman can deduce the results of the analysis. The output will be presented as a collection of charts and histograms

**System internal function design and implementation**

This is the third sub-stage and it is the most vital process of this research stage this will be the actualization of the theoretical framework as working executable code. It will involve defining and implementing the relevant functions that will automatically compute the results using the artificial intelligence (Machine Learning) algorithms.

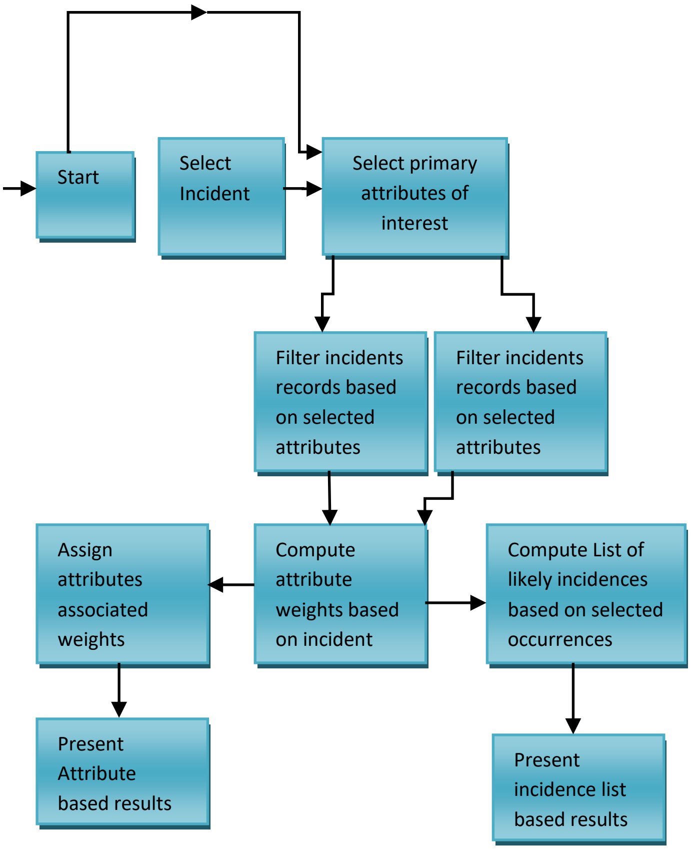

Start
Select Incident
Select primary attributes of interest
Filter incidents records based on selected attributes
Filter incidents records based on selected attributes
Assign attributes associated weights
Compute attribute weights based on incident
Compute List of likely incidences based on selected occurrences
Present Attribute based results
Present incidence list based results

Figure 1 System Framework internal control flow

**Data generation Prototype testing**

This stage will involve:

1. The actual generation of the simulation data for the model accuracy testing purposes.
2. The system prototype testing and comparison to externally predefined cases and conditions.
3. Observation of the results and comparison to the test cases for accurate reflection of the models validity.

**Results Documentation**

This is the final stage of the research project. This will involve the detailed documentation of the projects simulation and testing observations results and conclusions. This will be the final conclusion of the project as it is and the summarization of the results and observations will be documented for external analysis and possible implementation of the proposed research product.

## 4. KEY OUTPUTS

If the model can be borrowed and implemented effectively on a data set, it should improve:

- Intelligence analysis of crime data sets
- Strategic policy formulation using results of analysis above
- Intelligence led policing of masses

## 5. KEY ACHIEVEMENTS

The research has resulted to some noted achievements which include:

- The development of a quantifiable mathematical Model that can be used to design and implement a predictive model
- An Automatable framework that can be applied to any data set to bring out certain patterns and trends
- Extending the body of knowledge by introducing a novel hybrid data mining concept

## 6. KEY CHALLENGES

This research like any other has not been without its own challenges, both technical and resource wise as they include.

- Unavailability of actual real world data to compute real world events.
- Limited Research time as this research has proved to be wide and has a lot of factors that still need both interpretation and analysis
- Framework testing was limited due to the inadequacy of testing data required; occasioned by the sensitivity of the sector in which the intelligence analysis framework is designed to function in.

## 7. ASSUMPTIONS AND LIMITATIONS OF SCOPE

The completion of this research and development of this framework will present some challenges which are stated below.

- The framework will work the same if applied to any data set and prototype
- The accuracy of the framework model cannot be guaranteed since it has not been tested with real data.
- When this framework is adopted and used all policies within the law will allow its usage

## 8. CONCLUSION AND RECOMMENDATIONS

The development of this framework has and will enable the immediate and automated analysis of data sets. This will provide a justifiable basis of the policies that will be made by stakeholders within the relevant industry as it uses the factors that are of interest to events and individuals. If implemented it would considerably increase the capacity to quickly, easily and reliably analyse data. This will reduce the needless effort to develop half-baked rules and policies while ensuring that every major contributing attributes and entity sets are catered for.

## 9. REFERENCES

[1] Bruno Agard, Catherine Morency, Martin Trépanier. 2007. Mining Public transport user behavior from smart card data.

[2] Vikas Grover, Richard Adderley, Max Bramer. 2009. Review of Current Crime Prediction Techniques.

[3] Nassar, Khaled 2007. Data-Mining of State Transportation Agencies Projects Databases. 12.

[4] Tom M. Mitchell 1997. Machine Learning.

[5] Van der Veer, H.T. Roos, A. van der Zanden 2009. Data mining for intelligence led policing.

[6] Dale Dzemydiene, Raimundas Vaitkevicius, Ignas Dzemyda 2010. Pattern recognition based on statistics and structural equation models in multi-dimensional data warehouses of social behavioral data pg 4 -10.

[7] Han J, Kamber 2006. Data mining: concepts and techniques.

[8] Lee BS, Snapp RR, Musick R, Critchlow T. 2002 Metadata models for ad hoc queries on terabyte-scale scientific simulations.

# Assessment of Evidence Based Research on Internet Freedoms in Africa

Duncan Nyale
School of Computing & Mathematics
The Cooperative University of Kenya
Nairobi, Kenya

**Abstract**: The main objective of this work was to identify and map studies relevant to internet freedoms in Africa and analyze them to identify patterns and trends distribution in order to nurture public debate and for due consideration by researchers and policy makers within the region. The need to focus on internet freedoms in Africa is because of limited research contextualizing internet freedoms in this region and challenges with public availability and accessibility of the research evidence. The goal is to enhance learning on the research evidence available that could be relevant in advancing internet policies in Africa. The search for studies on internet freedoms was based on identified themes and indicators of internet freedoms. The study methodology was online desk research with full focus on empirical studies between 2013-2019.

**Keywords**: ICT: - Information and Communications Technology, IP: - Internet Protocol, USPF: - Universal Service Provision Fund, Fintech: - Financial Technology, ICT4D: - ICT for Development, DL: - Digital Literacy, SSA: - Sub-Saharan Africa

## 1. INTRODUCTION

Scoping of this assignment required conceptualization (or clarification) of internet freedoms. What is internet freedom or internet related freedoms? Internet policies? This research therefore included policies, laws/regulations, strategies and actions that affect control, direction, access, and use of the internet. The search for studies on internet freedoms was based on several identified themes and indicators of internet freedoms in each study. These included:

***Internet; digital rights and freedoms, freedom of expression on the internet***

*Studies on*:

- Enablers and/or barriers to access and use of the internet-costs, infrastructure including use Universal Service Provision Fund (USPF)
- Enablers and barriers to freedom of expression and enjoyment of human rights
- Hindrances/challenges to digital content and creative works
- Information control on the internet and hate speech
- Internet and social media disruptions and censorship-economic and social costs to online businesses; alert systems, browser traffic, internet user testing
- Internet governance-administration, management and control
- Digital human/civil rights violations

***Advocacy and civic engagement***

*Studies on:*

- Use of the internet in promotion of or demand for transparency, government financial accountability, democracy (in all its forms), e-participation, fight against corruption
- Enablers and/or barriers of use of internet in civic engagement; transparency and accountability of budget processes

***Intellectual property on the internet***

*Studies on:*

- Use of IP on the internet;
- IP theft and protection; copyrights and piracy on the internet
- Opportunities/challenges relating to IP issues on the internet

***Internet/digital/cyber security***

*Studies on*:

- Enablers and/or hindrances to privacy on the internet, who is affected most-minorities
- Surveillance and interception of communications-internet, social media monitoring and mobile phones
- Cyber violence, bullying and its impact on online businesses; effects on minorities

***Tech entrepreneurship and incubation hubs***

*Studies on*:

Enables and/or barriers to use of the internet in tech innovation, start-ups, acceleration, investment hubs; innovation and entrepreneurship ecosystems; e-commerce; capacity and skills development

- Impact/effectiveness/functioning of tech entrepreneurship/incubation hubs/platforms

***ICT4D and poverty reduction***

*Studies on*;

- Enables and/or barriers to the use of internet for health, education/digital literacy, fintech, poverty reduction;
- Impact of internet freedoms on digital literacy-what works, what does not
- Digital inclusion and exclusion-youth ICT capacity development; training

## 2. OBJECTIVES

The objectives of the study are as follows:

a) Mapping available research that will include collecting and cataloging existing, available and credible relevant research.

b) Analyze the collated data to identify patterns and classify the studies according to specified parameters.

c) Promote learning from existing research for stakeholders.

d) Promote the development of public policies from use of discrete pieces of research.

## 3. METHODOLOGY

The search for available studies on internet freedoms in Africa was conducted through online desk research focusing on empirical studies between 2013-2019. Search areas included:

- Online databases - Search Engines and Research Portals
- Institutional repositories - Research institutions, Universities, Advocacy organizations , Government agencies, International organizations and Professional associations/bodies
- Internet/ICT Regulatory Agencies - especially in SSA
- Social media - especially Twitter hash tags and Facebook posts of relevant internet forums

Sample search terms used to retrieve studies from the search areas included: Internet rights; digital rights; internet freedoms; digital freedoms; internet freedom of assembly; internet freedom; digital freedom of association; digital human; digital civil rights violations, internet violations; internet hate speech ,Internet governance**;** internet infrastructure management; internet administration; internet control; internet actors; content regulation; data protection; internet information control; fake news; misinformation and propaganda on the internet; internet and social media disruptions; internet shutdowns; internet censorship, IP regulation; content filtering; mandatory blocking of websites; end-user filtering; net neutrality; intermediary liability; network disruptions; Internet based entrepreneurship; e-commerce, internet-based start-ups; internet in tech innovation; internet investment hubs; innovation and entrepreneurship ecosystems; online businesses;

The articles retrieved were recorded and mapped according to the relevant criterion while factoring in crosscutting issues. The criterion used was:

### A. Themes

Studies addressing the following themes were identified and recorded:

- ***Internet Governance* -** "The development and application by governments, the private sector and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures, and programs that shape the evolution and use of the internet". Includes infrastructure management, internet administration and control, actors, content regulation and data protection, information control on the internet-fake news, misinformation and propaganda, internet and social media disruptions, shutdowns and censorship, IP regulation including creative works and copyrights, content filtering, mandatory blocking of websites, end-user filtering, net neutrality, intermediary liability, network disruptions
- ***Internet/Digital Rights and Freedom* -** Right to use the Internet and digital technologies in relation to freedom of assembly and association including through social networks and platform, digital human and civil rights violations, hate speech on the internet
- ***Internet for Democracy* -** internet in civic engagement/politics, transparency, accountability, participation, advocacy and civic engagement, transparency, accountability and fight against corruption, e-participation and e-politics
- ***Internet Safety* -** Online and cyber security, data protection and privacy online, the right to the protection of personal data concerning him or her, piracy, IP theft, cyber violence and bullying, unlawful surveillance, monitoring and interception of users' online communications and information by state or non-state actors including social media
- ***Internet Based Entrepreneurship*** - E-commerce, internet-based start-ups, Internet in tech innovation, start-ups, acceleration, investment hubs, Innovation and entrepreneurship ecosystems, online businesses, businesses through the internet and e-commerce
- ***Internet Access, Affordability and Use* -** access and use of online spaces, trends in fixed line and mobile internet, access to information/content, universal access and broadband strategies, affordability, cost of access, imposing taxes and fees on internet users etc.
- ***Digital Literacy and Skills* -** Refers to "as a set of basic skills required for working with digital media, information processing and retrieval" (UNESCO, 2011). DL involves the ability to use software or operate a digital device as well as a large variety of complex cognitive, sociological, and emotional skills that end-users need in order to function effectively in a digitally driven environment. Also refers to the awareness, attitude and the ability of an individual to use digital tools for communication, expression and social action in specific life situations. There are multiple forms of digital literacies: information literacy (digital content), computer literacy (hardware and software), media literacy (Text, sound, image, video, social), communication literacy (non-linear interaction) and technology literacy (Tools for life situations).
  - ▪ *Information literacy* can simply be distilled to refer to the ability to search, retrieve, manipulate, evaluate, synthesize and create digital content.
  - ▪ *Computer or ICT literacy* refers to the ability to operate digital hardware and software.
  - ▪ *Media literacy* encapsulates multiple streams of information and refers to the ability to interact with textual, sound, image, video and social medias
- ▪ *Communication literacy* refers to the ability to communicate in traditional and innovative mediums. This involves one-to-one communication in forms such as email, phone calls and short messages and

also in the one-to-many form, where an individual also broadcasts content across multiple mediums to reach a broad array of interested parties.

- *Technology literacy* refers to the ability to adopt various technologies to a particular life situation. Thus, knowing which tool to select is an important ability and being able to adapt the tool to a particular context is equally important. These skills are particularly needed in the IT sector and these involve the ability to create/maintain new products, services and digital technologies used in the modern economy.

**Note**: Crosscutting issues in each of the themes; Gender, youth, marginalized groups, people with disabilities were noted.

### B. Categories(Indicators of Internet Freedom)

Those that have implications on indicators of internet freedoms

- ***Trends*** in the themes identified above
- ***Enablers*** of internet freedoms (enhance people's use of the internet)
- ***Challenges*** to internet freedoms (makes it difficult to access and use internet)
- ***Impacts/Outcomes*** of programs and projects initiated to promote internet freedoms
- ***Solutions/Strategies*** to address internet freedoms such as pro-poor policies etc.

.

### C. Resources(Type of study)

- Book
- Book Chapter
- Thesis/Dissertation
- Research report
- Journal Article
- Conference proceedings
- Policy brief

### D. Region Study was Conducted

- ***Southern Africa***- Studies conducted in/from countries in this region
- ***East Africa*** - Studies conducted in/from countries in this region
- ***West Africa*** - Studies conducted in/from countries in this region
- ***Central Africa*** - Studies conducted in/from countries in this region
- ***North Africa*** - Studies conducted in/from countries in this region
- ***Africa*** – Studies conducted in/from regions that cut cross or are outside the directly specified

### E. Quality of the Studies

- ***High -*** Refereed or highly reviewed research evidence. This covers journals articles, book chapters, books. Have a high impact factor
- ***Medium -*** Have undergone some form of peer review. Includes conference Proceedings, Discussion Papers, Occasional Papers, Policy Briefs and Monographs, Theses/Dissertations (Postgraduate)
- ***Low -*** generally un-reviewed or minimal review. Includes research reports

## 4. DATA ANALYSIS AND PRESENTATION

Some 1040 articles were identified and analyzed as follows:

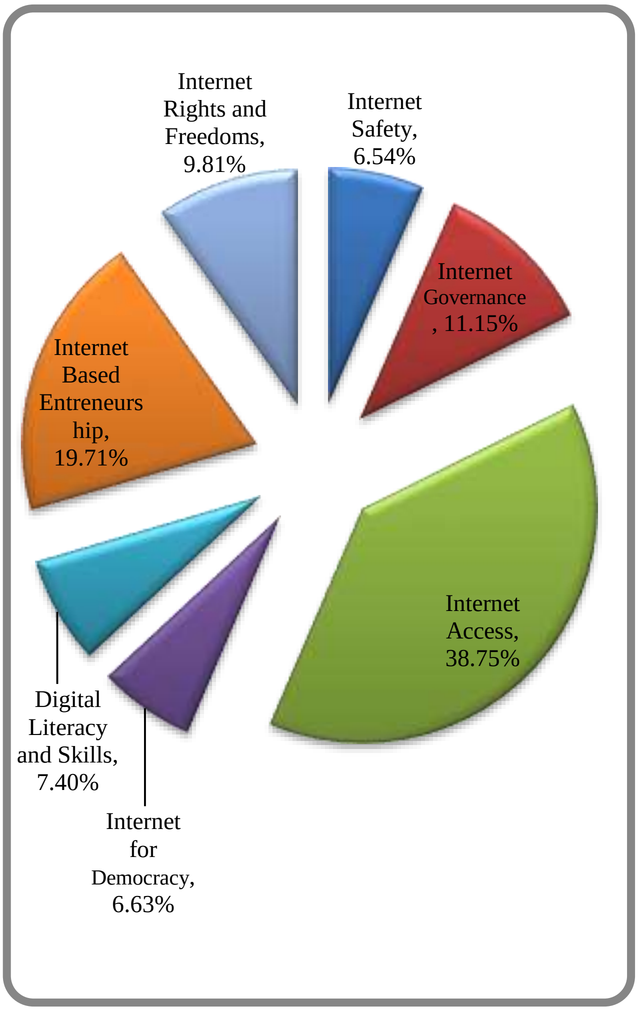


Figure 1. Studies Distribution by Theme

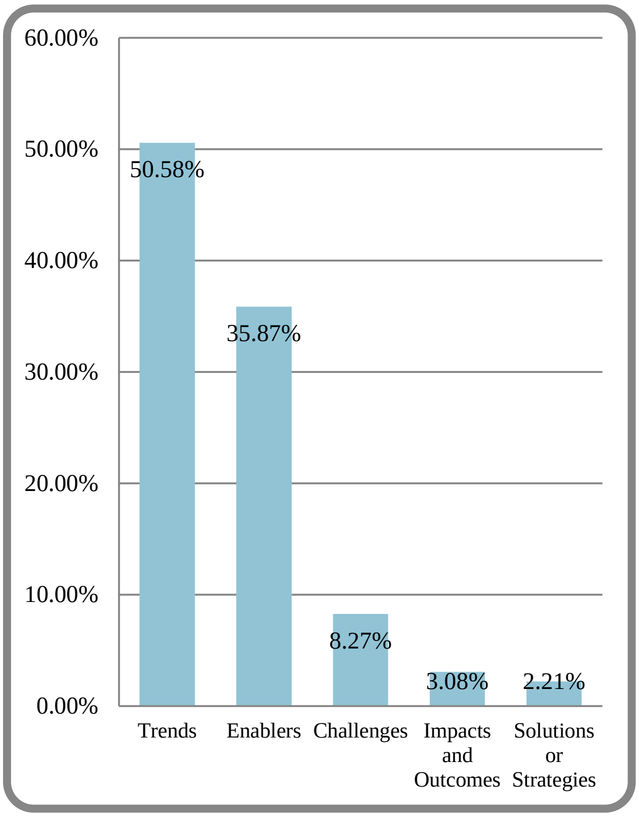


Figure 2. Studies Distribution by Categories (Indicators of Internet Freedom)

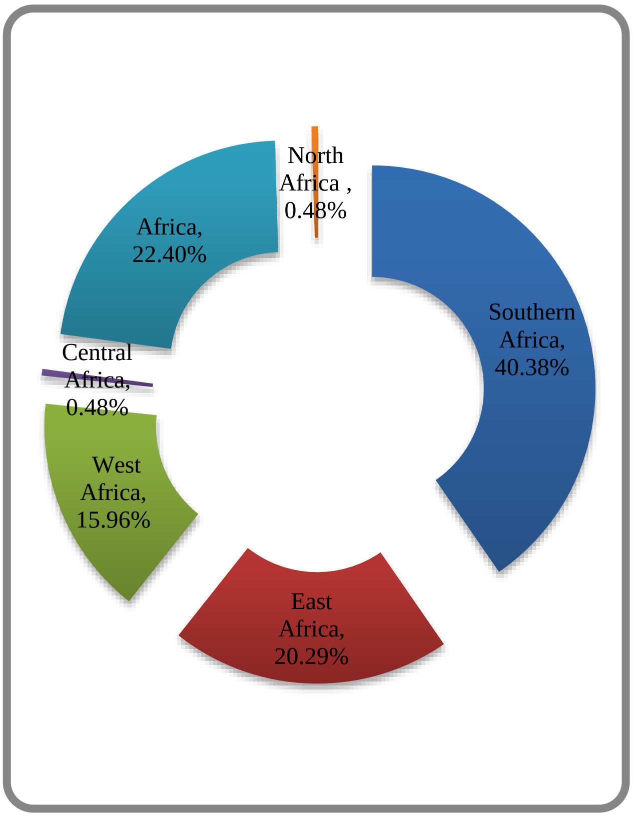


Figure 4. Studies Distribution by Region

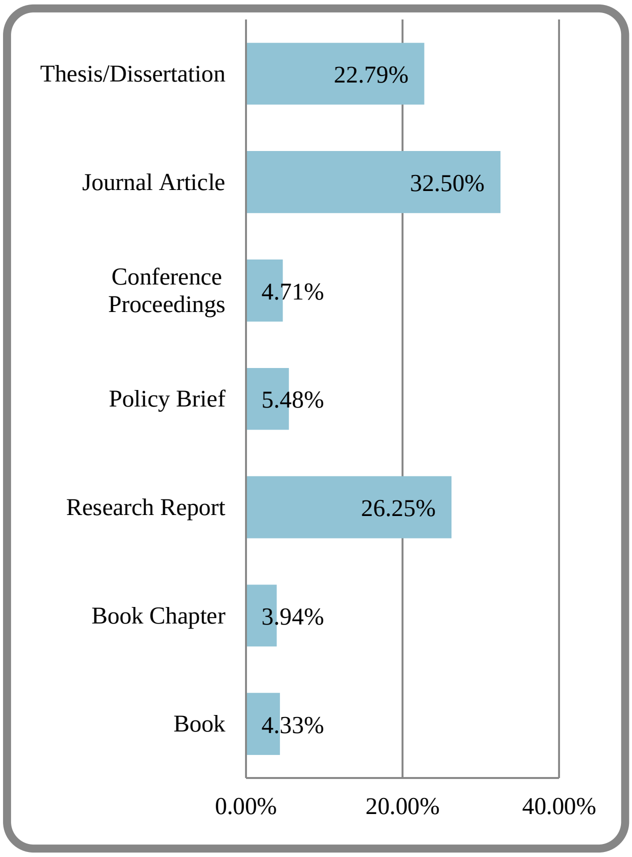


Figure 3. Studies Distribution by Resources (Type of study)

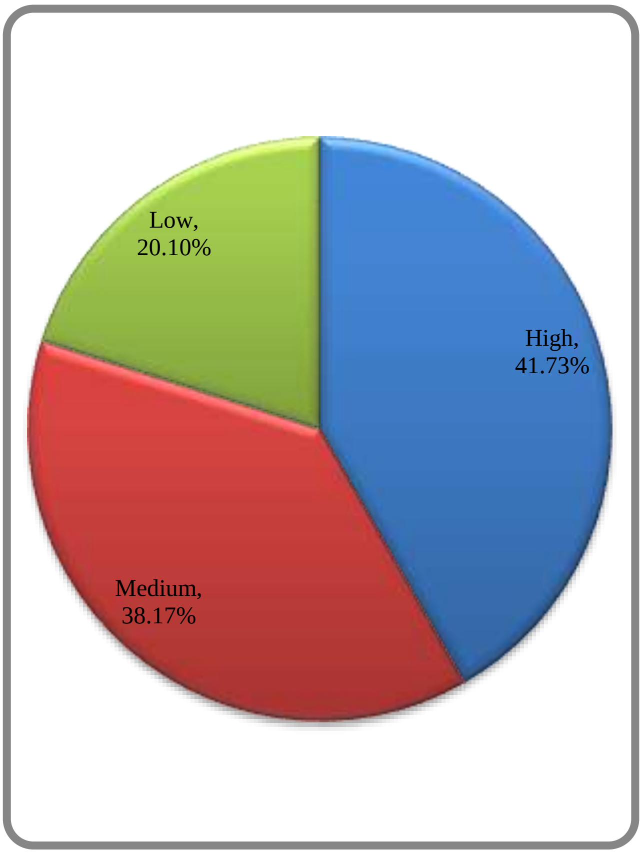


Figure 5. Studies Distribution by Quality

## 5. SUMMARY OF FINDINGS

Some of the key patterns that emerged from this research are:

- ***Themes***: Most of the studies are on Internet Access & Use (38.75%), while least are on Internet for democracy (6.63%) and Internet Safety (6.54%)
- ***Categories***: Most studies touched on Technological trends (50.58%), while least studied were Impact & Outcomes of application of internet technologies (3.08%) and Solutions and Strategies to internet technology challenges. (2.21%)
- ***Resources***: Majority of articles are Research reports (26.25%) and Thesis & Dissertations (22.79%) while least are Book Chapters (3.94%)
- ***Regions***: Most studies were in Southern Africa (40.38%) while the least studied regions were Central and North Africa (both at 0.48%)
- ***Quality of Studies***: Most studies were classified as High Quality (41.73%); followed by medium Quality (38.17%); and finally Low Quality studies (20.10%).

## 6. CONCLUSION

The conclusion of this research revealed that although there is considerable research on internet freedoms in Africa, it is heavily skewed of some themes, categories and regions. There is need to expand the scope to equitably cover the areas that are currently under researched. It is the hope of this researcher that the studies distribution herein can assist future researchers on which areas to lay focus on and thus improve the body of knowledge while ensuring acceptable distribution of research on all the research areas identified and analyzed.

## 7. RECOMMENDATIONS

- More research should be conducted in the regions currently under researched.
- Focus should be directed more to High Quality (Peer reviewed) articles.
- More research is required on Digital Literacy & skills, Internet safety and Internet for Democracy.

## 8. SAMPLE SOURCES OF EVIDENCE

[1] Google Scholar: http://scholar.google.com/

[2] Bielefeld Academic Search Engine (BASE): https://www.base-search.net/

[3] ScienceDirect: http://www.sciencedirect.com/

[4] WorldCat: https://www.worldcat.org/

[5] SpringerLink: https://link.springer.com/

[6] Google Books: https://books.google.com/

[7] CiteSeerx: https://citeseerx.ist.psu.edu/index

[8] Centre for Intellectual Property and Information Technology Law at Strathmore University (CIPIT): https://www.cipit.org/index.php/cipitresources

[9] University of Nairobi Digital Repository: http://erepository.uonbi.ac.ke/

[10] Collaboration on International ICT Policy for East and Southern Africa (CIPESA): https://cipesa.org/resources/

[11] Communication Authority of Kenya: http://www.ca.go.ke/index.php/research

[12] Uganda Communication Commission: http://www.ucc.co.ug/ and https://www.nita.go.ug/publications

[13] Social media especially Twitter hashtags and Facebook posts of relevant internet forums

[14] Combing through the hashtags #**InternetPolicyAfrica** #**InternetFreedomsAfric**a and identify any relevant ideas and sources of evidence.

[15] EbSCO: https://www.ebsco.com/products/research-databases

[16] JSTOR: https://www.jstor.org/

[17] Oxford Journals: https://academic.oup.com/journals/

[18] BudgIT: https://yourbudgit.com/data/publications/

[19] Taylor and Francis Online-Journals: https://www.tandfonline.com/openaccess/openjournals

[20] Google: https://www.google.com/

[21] Research ICT Africa: https://www.africaportal.org/content-partners/research-ict-africa/

[22] Paradigm Initiative: https://paradigmhq.org/reports/

[23] Oxford Internet Institute: https://www.oii.ox.ac.uk/research/

[24] UNESCO: https://unesdoc.unesco.org/ark:/48223/pf0000188700

[25] UNECA: https://repository.uneca.org/

[26] AfDB: https://www.afdb.org/en/knowledge

[27] Co-Creation Hub: https://cchubnigeria.com/focus/

[28] AU: https://au.int/en/resources/filter

[29] ITU: https://www.itu.int/en/publications/Pages/default.aspx

[30] The Alliance for Affordable Internet (A4AI): https://a4ai.org/research/

[31] Kenya ICT Action Network (KICTANet): https://www.kictanet.or.ke/?page_id=40115

[32] Freedom House: https://freedomhouse.org/policy-recommendations

[33] Open Access Gov: https://www.openaccessgovernment.org/category/open-access-news/research-innovation-news/

[34] The British Institute in Eastern Africa: https://www.biea.ac.uk/research/

[35] CODESRIA: https://codesria.kohalibrary.com/cgi-bin/koha/opac-main.pl

[36] Article 19 Eastern Africa: https://www.article19.org/law-and-policy/

[37] African Institute for Development Policy Research and Dialogue (AFIDEP): https://uia.org/journals

[38] The African Centre for Technology Studies (ACTS): https://www.acts-net.org/research/projects

[39] African Technology Policy Studies (ATPS): https://atpsnet.org/publications/

[40] Africa Freedom of Information Centre (AFIC): https://ifex.org/resources/

[41] The Africa Institute of South Africa (AISA): http://www.ai.org.za/research

[42] Botswana Institute for Development Policy and Analysis (BIDPA): https://bidpa.bw/publications/

[43] Institute of Statistical, Social, and Economic Research (ISSER): https://www.poverty-action.org/research

[44] World Bank: https://www.worldbank.org/en/research

International Journal of Computer Applications Technology and Research
Volume 9–Issue 05, 183-185, 2020, ISSN:-2319–8656

# Pneumonia Detection using X-Ray Images with Deep Learning

Hanumant Magar
MIT College of Engineering , Kothrud , Pune , India

Sanket J Patil
MIT College of Engineering , Kothrud , Pune , India

Sahil R Waykole
MIT College of Engineering , Kothrud , Pune , India

Satyam D Sandikar
MIT College of Engineering , Kothrud , Pune , India

Nikhil D Parakh
MIT College of Engineering , Kothrud , Pune , India

**Abstract:** This study proposes a Convolutional neural network model trained from scratch to classify and detect the presence of pneumonia from a collection of chest X-ray image samples. Unlike other methods that rely solely on transfer learning approaches or traditional handcrafted techniques to achieve a remarkable classification performance, we constructed a Convolutional neural network model from scratch to extract features from a given chest X-ray image and classify it to determine if a person is infected with pneumonia. This model could help mitigate the reliability and interpretability challenges often faced when dealing with medical imagery. Unlike other deep learning classification tasks with sufficient image repository, it is difficult to obtain a large amount of pneumonia dataset for this classification task; therefore, we deployed several data augmentation algorithms to improve the validation and classification accuracy of the CNN model and achieved remarkable validation accuracy. Our classification method uses convolutional neural networks for classifying the images and early diagnosis of Pneumonia. Our findings yield an accuracy of 85.73%, surpassing the previously top scoring accuracy of 78.73%.

*Keywords*—Pneumonia, x-ray imaging, early diagnosis, deep learning, automation

## I. INTRODUCTION

The risk of pneumonia is immense for many, especially in developing nations where billions face energy poverty and rely on polluting forms of energy. The WHO estimates that over 4 million premature deaths occur annually from household air pollution-related diseases including pneumonia .Over 150 million people get infected with pneumonia on an annual basis especially children under 5 years old . In such regions, the problem can be further aggravated due to the dearth of medical resources and personnel. For example, in Africa's 57 nations, a gap of 2.3 million doctors and nurses exists. For these populations, accurate and fast diagnosis means everything. It can guarantee timely access to treatment and save much needed time and money for those already experiencing poverty. Deep neural network models have conventionally been designed, and experiments were performed upon hem by human experts in a continuing trial-and-error method. This process demands enormous time, know-how, and resources. To overcome this problem, a novel but simple model is introduced to automatically perform optimal classification tasks with deep neural network architecture. The neural network architecture was specifically designed for pneu- monia image classification tasks. The proposed technique is based on the convolutional neural network algorithm, utilizing a set of neurons to convolve on a given image and extract relevant features from them. Demonstration of the efficacy of the proposed method with the minimization of the computational cost as the focal point was conducted and compared with the exiting state-of-the-art pneumonia classification networks. Pneumonia is an inflammatory response in the lung sacs called alveoli. Its often caused by bacteria, viruses, fungi and other microbes. As the germs reach the lung, white blood cells act against the germ and inflammation occurs in the sacs. Thus, alveoli get filled with pneumonia fluid and this fluid causes symptoms like coughing, trouble in breathing and fever. If the infection isnt acted upon during the early periods of the disease, pneumonia infection can spread throughout the body and result in the death of the individual, as a result of the inability to exchange gas in the lungs.

In recent times, CNN-motivated deep learning algorithms have become the standard choice for medical image classifications although the state-of-the-art CNN-based classification techniques pose similar fixated network ar- chitectures of the trial-and-error system which have been their designing principle. U-Net, SegNet, and Car- diacNet are some of the prominent architectures for medical image examination.

Models like evolutionary-based algorithms and reinforcement learning (RL) have been introduced to locate optimum network hyperparameters during training. However, these techniques are computationally expensive, gulping a ton of processing power. As an alternative, our study proposes a conceptually simple yet efficient network model to handle the pneumonia classification problem as shown in Figures 1 and 2.

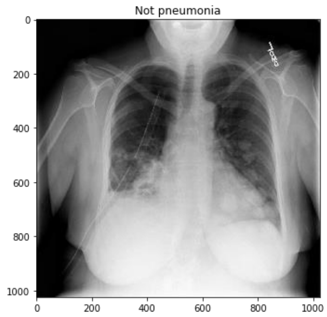


Fig.1

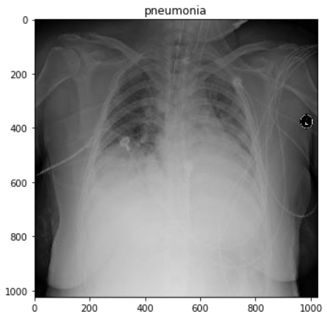


Fig.2

## II. RELATED WORK AND COMPARISON

There has also been previous studies done on the early detection of pneumonia. Among the various other methods used by different studies, this paper focuses merely on Pneumonia and its classification.

The experimentation was conducted in Koc University Artificial Intelligence Laboratory, Istanbul, Turkey. In their study they presented a novel method for classifying pneumonia existence in an x-ray image. They proposed a two-step image processing before training our deep learning model, in order for making the features of an x-ray image clearer and explicit for easing the classification process. They, then, executed a convolutional neural network followed by a residual neural network for the classification process.

Their experimentation was conducted with similar means to ours. They used a 3 layer convolutional network for feature map acquisition for image preprocessing. In our experiment we use 3 convolutional layers, yielding a more efficient and a computationally less costly training process. Our preprocessing methods are similar to real life applications, unlike statistical means that might be ineffective when wide range of data is present. Our proposed architecture yields an accuracy of 85.73%, while their study yielded an accuracy of 78.73%.

## III. MATERIALS

We present the detailed experiments and evaluation steps undertaken to test the effectiveness of the proposed model. Our experiments were based on a chest X-ray image dataset proposed in kaggle.com. We Used Keras open-source deep learning framework with tensorflow backend to build and train the convolutional neural network model. All experiments were run on a standard Laptop with an Nvidia GeForce GTX 1050Ti GPU card of 4 GB.

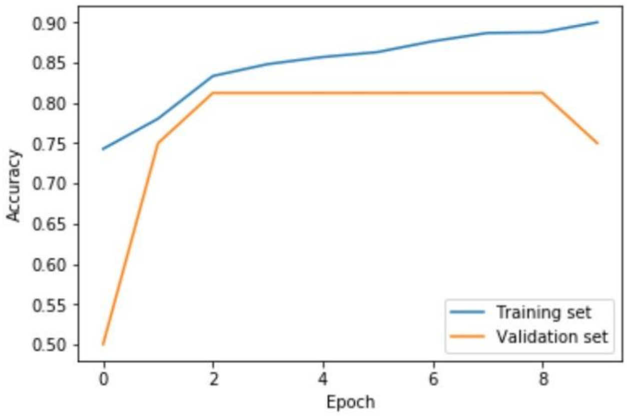
Dataset—The original dataset consists of three main folders (i.e., training, testing, and validation folders) and two subfolders containing pneumonia (P) and normal (N) chest X-ray images, respectively. A total of 5,856 X-ray images of anterior-posterior chests were carefully chosen from retrospective pediatric patients between 1 and 5 years old. The entire chest X-ray imaging was conducted as part of patients routine medical care. To balance the proportion of data assigned to the training and validation set, the original data category was modified. We rearranged the entire data into training and validation set only. A total of 3,722 images were allocated to the training set and 2,134 images were assigned to the validation set to improve validation accuracy.

## IV. MODEL

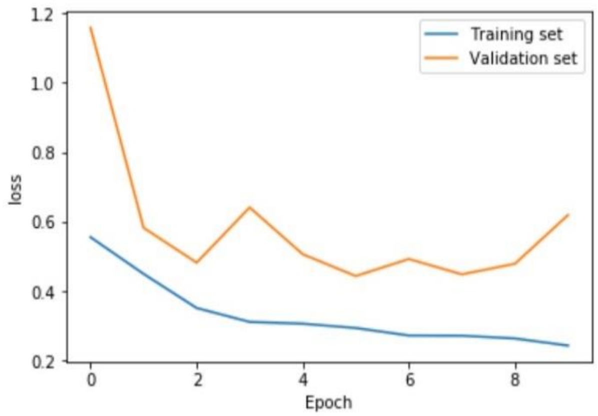
The model we build is a five layered convolutional network. The input image dimension is (100,100,1). The input images are generated using ImageDataGenerator class and augmentation techniques are used in order to shearing, shifting and zooming in image to match the input dimension. The model is trained in batches of 32 images. The activation function used is ReLu (rectified linear unit) for hidden layer and sigmoid for output layer. The loss function used is binary crossentropy because of binary output. The optimizer function used in the model is adam (adaptive gradient descent with momentum) optimizer. The evaluation metric we used for model is accuracy with stands for percentage of images correctly classified to total images used.

## V. RESULT

The accuracy curve is demonstrated in image below:



The Loss to Epoch graph is demonstrated in image as below:



## VI. MATHEMATICAL DETAILS BEHIND WORKING OF CONVOLUTIONAL NEURAL NETWORKS

CNN is primarily used to look for the patterns in an image. Feature selection in case of images which has hundreds and thousands of dimensions would be a painful process.

Feature selection happens all by itself during backpropagation over number of iterations. The input format of CNN is matrix ([[],[]]) and that of multilayer perceptron is tensor ([]).

Convolutional networks were inspired by biological processes in case of connectivity patterns of neurons. Unlike Multilayer Perceptron which uses fully connected layer, CNN uses different layers to detect patterns among images which are feed forwarded. Every image represent some pixels in simple terms.

We analyze the influence of nearby pixels on a particular pixel in an image by using a filter or a kernel.

Filters are tensors used track of spatial information. These filters learn to extract features like edge detection etc of objects in images in something called a convolutional layer. They help to filter out unnecessary or repetitive information during convolution operation.

There are multiple types of filters like high pass, low pass, gaussian etc which are used for different sort of operations. Then the convolution image matrix multiplies with filter matrix which is called Feature Map. Each filter is strided over the image using ot products between filter matrix and corresponding image pixel values. During the convolution operation size of image matrix decreases, if do not intend to do that then there is concept of padding. During padding redundant tensor of zero valued pixels is appended on edges of images, which preserve the size during convolutions. Next, we need to reduce the size of images, if they are too large. Pooling layers section would reduce the number of parameters when the images are too large.

Adding pooling layer then decrease the size of the image and hence decrease the complexity and computations.

Usually, an activation function ReLu is used in next layer. ReLU stands for Rectified Linear Unit for a non-linear operation.

The output is $f(\mathrm{x}) = \max(0,\mathrm{x})$.

The purpose of ReLu is to add non-linearity to the convolutional network. In usual cases, the real-world data want our network to learn non-linear patterns which is the purpose of activation functions. The final step is to flatten our matrix and feed the values to fully connected layer.

We need to train the model in the same way, we train other neural networks. Using the certain number of epochs and then backpropagate to update weights and calculate the loss.

## VII. CONCLUSION

In this study, we present a novel method for classifying an X-ray image on its possibility of exhibiting pneumonia in the early stags of the disease. We utilized a convolutional neural network approach for obtaining feature maps of the preprocessed X-ray images. Our method was convolutional neural network which yielded a high accuracy. Our most accurate experimentation model classifies the images with a 85.73% accuracy. Overall, we target the current drawback in medical diagnosis of pneumonia by the human high, and propose an alternative and more accurate way of diagnosing the disease with automation. Moreover, we target the limits caused by the gray scale of x-ray imaging, preventing the early diagnosis of the disease. Our study presents an efficient algorithm with a high performance for this classification task and can be improved through object detection algorithms for extracting the region with pneumonia.

## VIII.REFERENCES

1. Deniz Yagmur Urey, Can Jozef Saul, and Can Doruk Taktakoglu
Robert College of Istanbul, Istanbul, Turkey
Koc University Artificial Intelligence Laboratory, Istanbul, Turkey
Early Diagnosis of Pneumonia with Deep Learning.

2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

3. CheXNet:Radioogist-Level Pnemonia Detection on Chest -Ray with Deep Learning. https://arxiv.org/abs/1711.05225v3