Estimating the Survivability of Patients Using Prediction Markets and Cox Hazard Proportion Regression

Henry Asante Antwi\textsuperscript{ab*}, Zhou Lulin\textsuperscript{ab}, Tehzeeb Mustafa\textsuperscript{ab}, Numair Nisar\textsuperscript{ab}

\textsuperscript{a}School of Management, Jiangsu University, Zhenjiang 212013, PR China.
\textsuperscript{b}Institute of Healthcare Management, Jiangsu University, Zhenjiang 212013, PR China.

Corresponding Author: lulinzhou@yahoo.com

ABSTRACT

Available literature shows that a large group's aggregated answers to questions involving quantity estimation, general world knowledge, and spatial reasoning has generally been found to be as good as, and often better than, the answer given by any of the individuals within the group. This has inspired the development of new models of crowd wisdom techniques such as prediction markets. To find out whether prediction market or crowd wisdom tools can help in making accurate prognosis, one medical case was chosen which is how well experienced doctors can predict the survivability of patients with gastric cancer. The results are compared with the outcome of survivability forecast using the Cox hazard proportion regression model and the artificial neural networks. The ANN accurately forecasted 31% of patients to survive whiles 33% will not. On the other hand, the Cox Hazard Model accurately predicted 29% of the patients to survive whiles 31% will not. Finally, the PM market predicted 31% of the patients to survive whiles 31 percent will not survive. On the whole the prediction accuracy of the ANN was 64% whiles that of the CPH and the Prediction Market were 60% and 62% respectively. This implies that that whiles the ANN defeated the PM Model in predicting accuracy of survivability of patients with gastric cancer, it outperformed the Cox Hazard Model by 32 percentage points.

Keywords: Survivability, Prediction, Cox Hazard, Comparison, Prediction Markets, Neural Networks.

INTRODUCTION

According to Kittur (2007), a large group's aggregated answers to questions involving quantity estimation, general world knowledge, and spatial reasoning has generally been found to be as good as, and often better than, the answer given by any of the individuals within the group. An explanation for this phenomenon is that there is idiosyncratic noise associated with each individual judgment, and taking the average over a large number of responses will go some way toward cancelling the effect of this noise (Kitzinger & Kitzinger, 2015). Prediction market is one of the many crowd wisdom tools that have been developed over the years as a basis for forecasting to aid organizational decision making. In a prediction market, experts are asked to forecast the possible occurrence of an event. The results are then compared with conventional prediction market models to determine their prediction models (Duru, et al, 2012). In the healthcare sector prediction market has been used by many researchers in both clinical and non-clinical decision making with conflicting outcomes (Chatfield, 2013). For example, in 1976, Chang et al (2007) randomly selected a sample of 65 general practitioners and 78 medical and surgical gastroenterologists to predict the likely current state of a cohort of 227 patients first diagnosed with duodenal ulcer in 1963 in hospitals and general practice. This was after the experts had extensively reviewed the medical profile of each patient. At the time the actual state of the 227 patients showed that 50 patients had died, 57 had been medically treated with no symptoms, 44 had mild symptoms, and 34 had been treated surgically while 19 of them had more severe symptoms. The remaining 12 had
emigrated. The study noted that cases that had been diagnosed in hospitals had a more severe prognosis than those diagnosed in general practice. The individual prediction deviation of the experts was very wide showing that individual prediction estimate less reliable (Wolters, 2008).

However, the mean prediction level by all doctors differed marginally from the actual estimates suggesting the reliability of collective experience of the medical profession. The study also found out that the general practitioners, surgeons and physicians showed insignificant systematic differences; a reflection of the differences in the types of patients they treat. However, the healthcare setting involves both clinical and non-clinical decision making all of which require predictive competence of experts (Wolters and Zitzewitz, 2014). One of such areas is the arrival of patients in the hospital in general and departments in particular. For example predicting the accurate number of patients arriving at the emergency department is very critical for healthcare planning. Developing and testing new models to aid decision making is necessary in alleviating the challenges associated with conventional statistical or predictive models (Caporale & Gil-Alana, 2013). This chapter explores the extent to which prediction market as crowd wisdom techniques are useful in predicting or forecasting health service outcomes. Even though there are ways patients judges the quality of services they receive from the hospital, the degree to which a therapy or medical process they undergo makes them well is the ultimate measure of satisfaction. That is why the doctor’s prognosis must be accurate. To find out whether prediction market or crowd wisdom tools can help in making accurate prognosis, one medical case was chosen which is how well experienced doctors can predict the survivability of patients with gastric cancer. The results are compared with the outcome of survivability forecast using the Cox hazard proportion regression model and the artificial neural networks. This helps to determine which of the method accurately predicts patient survivability.

RELATED WORKS
Judgmental forecasting models seeks to incorporate intuitive judgment, personal or group opinions and subjective probabilities to support decision making when historical data is not enough. To date the most common judgmental forecasting models includes composite forecast, Cooke’s method, Delphi method, prediction market, simulation, probabilistic models, ensemble models, scenario building, statistical survey and technology forecasting. Indeed early accounts of the development of modern medicine suggest that the clinical skills, scientific competence and doctors’ judgment were the main impetus for treatment decision, diagnosis, prognosis, therapy assessment and medical progress hence the quality of services delivered by the clinician(Uchino, Bellomo, Goldsmith, Bates, & Ronco, 2016).

However, depending solely on the judgment of a single clinician or isolated cases of unrelated hospital experiences to make healthcare decision can affect the quality of decision made and this can invariable affect the quality of services provided to the patient. This is because some researchers believe that an individual clinical judgment can be notoriously fallacious and may appear as an irrational and unfathomable black box with little transparency (Akins, Tolson, & Cole, 2005; Ohno-Machado, 2017). This explains why even though the past decade has seen the emergence of several new investigations and theories about applying judgment in clinical and related healthcare decisions, most of them have been restricted to its role in communication, late stage diagnosis, prognosis and other medical decision scenarios with the greatest caution about its susceptibility to error and bias as opposed to their validity, potential competence, reliability and theextent to which it can be optimized and professionalized to enhance effective forecasting and service quality delivery(Bederman et al., 2017; Ohi et al., 1987).

Anderson et al (2016) reiterate this by explaining that with the rise of contemporary medical research, the reputation of clinician judgment underwent significant reform in the last century as its fallacious aspects were increasingly emphasized relative to the evidence based options. Critics of clinical judgment as a healthcare decision making tool for example presumes that it cannot go beyond a simple post-hoc-ergo-propter-hoc, but can at best achieve simple, intuitive, low-quality correlational statistics for healthcare decision making (Bornstein & Emler, 2017; Mushlin & Appel, 1977). Coupled with an increasing numbers of judgmental errors on the part of doctors, a primary mission was initiated ‘to guard against any use of judgment in healthcare decision making(Christensen & Elstein, 1991; Kittur, Chi, Pendleton, Suh, & Mytkowicz, 2016) while emphasis moved to the exploration and use of clinical trials and statistical models.

Since the 1960s the “anti-guessing” theory of Evidence Based Medicine (EMB) currently practiced globally by clinicians and allied health services has dominated medical practice and
associated healthcare decision making following series of publications by Alvan Feinstein, Archie Cochrane, John Wennberg, David Eddy, David Sackett etc (Bradbury et al., 2017; Jones, 2017; Smart, Blake, Staines, & Doody, 2017). Jones (2017) even claim that in main stream hospital management that is traditionally a non-clinical area of healthcare management, evidence of importation of evidence based medicine to decision making process is dominant since most healthcare facilities are managed by clinicians and other personal with evidence based training and orientation.

As an optimized healthcare decision making approach, EBM emphasizes evidence from well designed and executed research as the fulcrum of all medical and related decisions. Even though all medicine based sciences have some degree of empirical validation, EMB goes further by classifying evidence by its epistemological strength and recommends only the strongest types (coming from meta-analyses, systematic reviews and randomized controlled trials with the support of statistical tools) (Diaz-Aviles, Stewart, Velasco, Denecke, & Nejdl, 2017).

While the important place and role of EBM in contemporary medical practice is strongly represented in modern healthcare literature, it also has its fair share of criticisms. For example (Soll & Larrick, 2017) criticizes EMB for its restricted process of evidence collection and approval. They contend that “EMB sometimes suffer from a ‘Central Control’ phenomenon as a few chosen experts are tasked with the responsibility of digging out evidence, then instruct others on how to interpret and utilize the evidence”.

Moreover Webster & Ksiazek (2017) argues that the quantitative results produced by EBM research especially from randomized controlled trials may be irrelevant for some treatment situations while racial minorities and people with co-morbid diseases which are usually under researched may limit the generalizability of randomly controlled trials. Zhu et al. (2013) reports of disparities between treatments effectiveness reported from randomized controlled trials with those achieved in routine clinical practice and population based research which EBM champions may not apply on a patient by patient basis. Thus in most instances, the knowledge acquired from clinical research studies to design evidence based standards fails to directly address clinical questions regarding what is best for the patient at hand.

Within the last decade, medical forecasting literature has seen significant attempt to revisit the role of clinician (doctors) judgment in medical decision making as a complement of EBM due to its practical limitations. Shi, Lee, et al (2017) stresses that the grand attempt to discredit the use of personal judgment by clinicians in the 1960s was not based on systematic investigations but on selectively procured sample of judgmental error or sometimes anecdotal examples of error and naivety on the general low esteem of personal cognition in the times of neo-positivist (Chen et al., 2016; Graefe & Armstrong, 2017; Smart et al., 2017) and fallibilist (Chen et al., 2016; Tepper, Dejong, Willkerson, & Brannon, 1995; Wei et al., 2013) epistemologies. (Liu et al., 2013) and other “radical” advocates of clinical judgments emphasize that the experiences of different expert (clinicians) can complement EBM in specific medical decision scenarios such as when treating new illness with limited statistical data, prognosis of survivability of a particular disease (Lowthian et al., 2017), when there are few records of patient data with given symptoms etc. In that case making available the judgment or experiences of physicians who have encountered several such cases during years of practice can provide valuable additional information for decision making.

Ansari et al (2013) affirm this by proposing that in some sense experts are human measuring instruments. Just as a sensor can measure a patient’s blood pressure, temperature etc, the experience of a medical expert can supplement these measurements in diagnosis and prognosis. This argument is reasonable to the extent that experienced and competent professionals rely on both explicit factual evidence and their tacit knowledge before making any decision (Bollschweiler et al., 2014; Chi, Street, & Wolberg, 2016; Colak, Colak, Kocatiurk, Sağiroğlu, & Baruçu, 2016; Delen, Walker, & Kadam, 2005; Peng & Peng, 2016). Any competent practitioner worth his or her profession is disposed to make several judgments of which the specific or adequate criteria cannot be easily expressed and equally displays skills whose rules and procedures cannot to be explicitly stated. In this case he or she depends on tacit recognitions, judgments, and skilful performances to draw conclusions which are mostly accurate (Bassi et al, 2016; Gohari, Biglarian, Bakhshi, & Pourhoseingholi, 2017; Jerez-Aragónés, Gómez-Ruiz, Ramos-Jiménez, Muñoz-Pérez, & Alba-Conejo, 2013; Zhu et al., 2013). Thus “there is a clamour to represent individual variety in medical prognosis and corresponding decision making through alternative but accurate prediction approaches and should be provided a platform for presentation. However a more conservative view in the clamour to represent clinical judgment in the medical decision
making process has emerged to help control potential clinician abuse. (Wang et al., 2017) rather advocates for what they call a “cybernetic variety” that deemphasizes individual doctor’s judgment and rather proposes the creation of a “pool of experiences” from which clinicians can draw experiential information when faced with context specific medical dilemma. In this way using “crowd wisdom” approach instead of “individual wisdom” is presented as a more credible option to complement EBM and help gather all available knowledge, experiences, possible alternatives or bits of information from experts together to treat specific healthcare cases (Mariani et al., 1997; Maroco et al., 2017).

With the advent of swarm intelligences, the healthcare sector has been one of the many notable areas where experiential knowledge of both clinicians and non-clinicians is being harnessed and optimized in a swarm environment to improve demand forecasting and service quality etc due to the uniqueness of the service industry. Beside swarm intelligence, many other crowd wisdom techniques (Delphi, prediction market, scenario building etc) have been used to forecast healthcare demand to ensure accurate forecasting of patient conditions in order to provide them with high quality services (Berkman, et al, 2014).

The relevance of harnessing crowd wisdom or judgment techniques to support effective health service quality delivery is well articulated by Tandberg and Qualls (2014) who explains that doctors for example develop skills to make effective medical judgment or treatment quality based on the experiences from practice and knowledge shared with comrades, critical analysis, continuous research and ongoing professional development. This extends to all medical areas including diagnosis, therapy, prognosis, communication and other non-clinical decision making.

Burke et al as stated in Turner, et al (2017) espouses the innate wisdom of the crowd as opposed to individuals in the story of “cleaning the crystal ball”. This story discusses the challenges of prediction using the old game of estimating the number of jelly beans in a jar. In a 1987 study conducted by Professor Jack Treynor, 56 students were asked to provide estimates of how many jelly beans were in a jar. The mean guess of the students was 871, representing a 97.6% level of accuracy, with only one of the 56 estimates getting closer to the actual value of 850 see (Grainger & Griffiths, 1994; Kors & van Bemmel, 1989). In support of Treynor’s work, a similar study conducted by the researcher, again sampling estimations from 56 students showed a similar level of accuracy of 98.7%.

According to Hickey & Roberts (2014) using “crowd wisdom” in medical decision making is driven and embodied by Ashby’s Law which is applicable in many forms. “Ashby’s Law” stipulates that the minimum amount of information needed to give an accurate answer is exactly the amount needed to specify the problem. This is interpreted as; if the question has lot of variety the answer too will have the same amount of variety. A complicated question will obviously not have a simple answer either. In clinical decision, management of a complex fracture in patient with multiple co-morbidities in a resourcefully challenged situation cannot be resolved by ‘Cookbook’ approach presented by evidence based medicine. Thus if we need an answer to a complex situation, more information will be needed on a large scale and pooling the ‘wisdom of the medical crowd’ will be more effective than a controlled approach (Jones, 2014).

This study attempts to determine the degree to which wisdom of the crowd tools such as “prediction market” is applicable within the healthcare market based on the outcome of two studies. The first is a non-clinical study of forecasting patient arrival at the emergency department in two referral hospitals in China (Jiangsu and Guanxi Provinces). The objective is to provide evidence for the effective use of wisdom of the crowd theory in general and prediction market techniques in particular to forecast health service demand and how accurately doctors can use it make prognosis (predicting the outcome of a medical case) to enhance healthcare quality delivery. The results are then compared with the outcome of similar predictions made using existing medical forecasting techniques. Firstly the prediction market technique is compared with the exponential smoothing, seasonal autoregressive moving averages, time series regression and artificial neural network to predict patient flow in the emergency department of the affiliated hospital of the Jiangsu University and the affiliated hospital of the Guilin Medical University in the Guanxi province.

Secondly the outcome of the prediction market technique is also compared with the prediction accuracy of using the Cox hazard proportion regression model and artificial neural networks to forecast the survivability of patients with gastric cancer in China after reviewing and modeling their histological information. If successful the prediction market technique will not only become useful in the case of clinical decision making but also provide
more accurate forecasts models than traditional methods of healthcare demand forecasting and can assist in bed space organisation, staff planning and cost reduction mechanism for hospitals rather than using alternative demand estimating models some of which are complex and time consuming.

MATERIALS AND METHODS

Data to set up the prediction market, run the Cox hazard proportion regression and the artificial neural networks was sampled from the histological records of a cohort of 150 patients diagnosed of different stages of gastric cancer in selected hospitals in China. Each patient folder contained three set of information useful for the purposes of the prediction market, Cox Hazard Proportion Model and Artificial Neural Network. Firstly, the patient folders contained data about the severity of each patient condition classified using the American Joint Committee on Cancer TNM classification (7). From this, it was determined that before the surgery fifty six (56) patients suffered from Stage I gastric cancer, while eighty six (86) suffered from Stage II gastric cancer. Finally one hundred and sixteen (116) patients suffered from stage III cancer with the remaining thirty one (31) of them suffering from stage IV gastric cancer. The second information in the folder indicated the type of surgery each of them underwent. Consistent with their respective diagnosis, one hundred and sixty eight (168) patients representing 58.1 percent of the patients had undergone palliative gastrectomy while forty one (41) of them underwent radical subtotal gastrectomy. The remaining seventy six (6) patients went through radical total gastrectomy. The third set of information in each patient folder was a record of their patient folder for the purposes of the analysis.

Regard the Prediction Market model, the study randomly selected a sample of 40 international post graduate medical students from selected medical universities in China with specialization in cancer studies. They were asked to predict the survivability of a cohort of 150 patients first diagnosed and operated on gastric cancer in general practice. Each doctor had the opportunity to evaluate the detailed case summaries as recommended by (51). The one month experimental prediction market was opened at 12:30pm on the 13th of September 2016 and closed at the same time on the 12th of October 2016 and doctors could submit their prediction at any time within that period. A total of 210 estimate of survivability of 149 individual patients from 50 doctors were received on the 12th of October 2016. Prior to the study written consent forms, which were approved by the Ethics Committee was sent to the participants to sign before participating in the study. The consent form affirmed the participant’s voluntariness to participate in the research based on understanding of the objectives and process of study. They affirmed their consent for anonymity of the research, the confidential holding of the data up to a reasonable time after the publication of the outcome of the research findings and other standardized requirements by the Ethics Board. Again as a measure to incentivize the staff to voluntarily and assiduously participate in the process (as suggested by(52, 53), the top ten accurate forecasters were promised a token monetary reward of ¥500 for prediction accuracy. The study was conducted with optional anonymity in order to comply with ethical issues and gathered the following characteristics for each participant: age, occupation, department, number of years worked in the respective hospitals and number of years worked in their field whether in
or outside of their respective hospitals. Actual data on the medical condition of the 149 individual patients whose data was collected were updated in our system from the respective hospitals on the 12th of October 2016; being exactly a year after predictions from Senior doctors were closed. This actual data, as stated by Rajakovich & Vladimirov 2008, is expected to compare the accuracy of the participants’ estimations. At the time the actual state of the 150 patients showed that 30 patients had died, 37 had been medically treated with no symptoms, 44 had mild symptoms, and 34 had been treated surgically while 5 of them had more severe symptoms.

Cox Hazard Proportion and Neural Network Models

This section evaluates the prognostic accuracy of using neural networks and Cox hazard proportion models to predict the survivability of patients after 12 months. The data for analysis is the historical data of patients with gastric cancer collected from selected hospitals in China. Consistent with other prediction earlier works the main variables of interest are the BMI, Stage of disease, Peritoneal Dissemination, Histological Grade, Histological Type, Radicality of surgery, CEA level and CA19-9 level. The main demographic variables that were added include the gender and the age in order to determine the type of people that survive most of not. To these social conditions and location were added. This is consistent with the observation of Garcia and Chan (2017) that extraneous factors have a significant influence on survivability of patients.

Preliminary Analysis

Table 3: Reliability and Validity Test

The information in table 3 gives the output of reliability test of internal consistency of the variables used in constructing the Cox hazard and artificial neural networks. As observed by Pallant (2016), the threshold of reliability must be 0.9 below which the construct ought to be dropped. Significantly, all the constructs that used in the model exceeded the designated threshold of 0.9 as observed in the table.

Test of Gaussian (Normality) Distribution

Table 4: Output of aggregated normality test for independent variables

This test was conducted to whether the data was normally distributed or not. The normality of data is important in the analysis because it helps to determine the specific form of inferential analysis (parametric and non parametric) that should be conducted. The Kolmogorov-Smirnov and the Shapiro Wilk tests (Franke and Hardle, 2015) were thus conducted as and the results are displayed in table 10. Generally, the test results of all the 12 variables in the analysis indicate has a p-value more than 0.5 significant levels. This implies that each of the data is normality distributed hence a parametric analysis can be used (Pallant, 2016)

Test of Multicolinearity

Since the study involves the use of regression analysis of the text of causes and effect relationship, there was the need to determine the existence of multicolinearity was conducted before the regression analysis. Multicolinearity occurs where there is a significant correlation between the independent variables. This reduces their independence in prediction and can undermine the predictive power and accuracy of the model. Different measures can be used to detect possible multicolinearity between the independent variables but this study adopted two models namely the intercorrelation matrix as well as the Variance Inflation Factor (VIF).
Table 5: Results of Correlation Matrix of Independent Variables

Table 5 shows the outcome of Pearson Moment Correlation (parametric) correlation matrix to detect multicollinearity. The tests suggest that there is no problem of multicollinearity. As a rule of thumb, Fields explains that multicollinearity exist when the correlation between two independent variables is 0.5 or higher. In the model the highest correlation is .405 and that is an indication of the fact that all the variables are independent enough to be used for the regression and test. This information is affirmed by the outcome of the Variance Inflation Factor test in table 6 (appendix)

By examining VIF in table 18, all independent variables have VIF less than 10 (the maximum VIF is 1.741 for the independent variables. This implies that all the variables are good factors for predicting survivability of patients using both Cox hazard and artificial neural network models.

Figure 1 provides a diagrammatic representation of the artificial neural network model developed to predict the survivability of patients with gastric cancer based on the histological records obtained from the patient folder. The SPSS Version 20 was used for analysis. To validate the data, the Kaplan-Meier and log rank tests were performed. A more advanced patient survivability test was performed with the Artificial Neural Networks (ANN) and the Cox Hazard Proportion Regression. The patients were divided into two sets with one set (173 patients) treated as training set to construct the models and the remaining 116 patients used as testing sets to validate the data. Overall the structure of the neural networks included a three layer back propagation made up of eleven (13) input layer nodes, six (5) hidden layer nodes and one (1) output layer node was constructed as demonstrated in figure 13. A dichotomous or binary response to the status of patients (deceased or survived), a sigmoid function was used as the activation function in the hidden layer of output. The sigmoid basis function is a mathematical function with an "S" shape (sigmoid curve) and relates to the reasoning of the brain. The structure of the sigmoid basis function helps the computational believers and is often referred to in special case of logistic function (nonlinear).

Yet Al-Qahtani and Crone(2013) explains that the sigmoid function exist in different forms and have
been widely applied in different computation and operational contexts. For example sigmoid basis function such as the logistic and hyperbolic tangent weight functions are used as activation functions of neurons. Similarly, there is a profuse use of sigmoid functions in statistics such as integrals and logistic distribution, normal distribution, and student's probability density functions (Al-Quhtani and Crone, 2013). Thus sigmoid basis functions gives nonlinear effect for large input value and at the same time provides nonlinear effect at small input value (Amiri, et al, 2013). Following, a back propagation learning algorithm learning rate of 0.05 and a momentum of 0.9 was used for the net training. The threshold to discontinue the learning process in the training set was fixed at a mean square error of 0.0001. In the case of the Cox Hazard Proportion model, a backward selection method was used to fit it at a significant entry level of 0.01 and a significant removal level of 0.015. The modeling process was divided into two stages. Firstly 60% of the patients were separated into a training set while the remaining 40 percent were separated as a subset. To compare the survival distribution of the training and the subset samples, a log rank (Mantel-Cox test) was carried out. The efficiency of the Mantel-Cox test in validating the extent to which a control group compares with an experimental group are explored by many researchers such as Garcia and Chan (2017) hence the appropriateness of its use. The outcome of the test revealed that there was no significant difference in the survival curves using the training and the testing subsets (p = 0.650 > 0.5). Secondly based on the validation set, the Cox Hazard Proportion and the Artificial Neural Network models were employed to determine the risk factors and the results are represented in table 13.

**Table 7: Results of CPH Prognostic Factors**

<table>
<thead>
<tr>
<th>Results of Prognostic Factors</th>
<th>CPH</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Peritoneal dissemination</td>
<td>0.0039</td>
<td></td>
</tr>
<tr>
<td>Radical surgery</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td>Histological grade</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>CEA</td>
<td>0.295</td>
<td></td>
</tr>
<tr>
<td>CA19-9</td>
<td>0.3929</td>
<td></td>
</tr>
<tr>
<td>Ascites</td>
<td>0.7224</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>0.7428</td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>0.7972</td>
<td></td>
</tr>
</tbody>
</table>

Table 7 shows the results of the ordered factors of the prognostic variables using the Cox hazard proportion regression model. This indicates the effect of each of the prognostic factors on patient survivability. The results show that the stage of stage is the most significant predictor of the survivability of the patient (p=0.001) and this is followed by the effect of the peritoneal dissemination on the survivability of the patient (0.0039). Further, the results indicates that the radicality of the surgery is third most significant predictor of the survivability of the patient (0.0216) while the Body Mass Index (BMI) is the fourth most significant predictor of the survivability of the patient (0.0902). The data shows that Age and Histological grade are the fifth and sixth most significant predictor of patient’s survivability with p- values of 0.0313 and 0.0128 respectively. Moreover, the level of CEA is the next most significant factor affecting patient survivability (0.2695) and this is followed by the level of CA19-9 (0.3929). On the other hand, the Ascites significantly influences patient survivability (0.7224) as closely as Gender (0.7428) and the Histological type (0.7972). Thus the information gives an indication that almost all of the predictor variables are strong predictors of survivability with highlight significant p values.

**Table 8: Results of Normalized Importance of ANN Factors**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.23377</td>
</tr>
<tr>
<td>Gender</td>
<td>0.21534</td>
</tr>
<tr>
<td>Radicality</td>
<td>0.11155</td>
</tr>
<tr>
<td>Peritoneal Dissemination</td>
<td>0.07178</td>
</tr>
<tr>
<td>Histological grade</td>
<td>0.06499</td>
</tr>
<tr>
<td>CEA</td>
<td>0.04171</td>
</tr>
<tr>
<td>Age</td>
<td>0.04074</td>
</tr>
<tr>
<td>Gender</td>
<td>0.03104</td>
</tr>
<tr>
<td>Ascites</td>
<td>0.02813</td>
</tr>
</tbody>
</table>

Table 8 on the other hand focuses on the artificial neural networks by presenting the normalized importance of each of the predictor variables. The study found out that the most predominant variable is the stage of disease (0.23377) whereas the radicality of the surgery follows (0.21534). Further, the level of CA19-9 is determined to be the next important predictors (0.11155) whiles the peritoneal dissemination follows with a normalized importance value of 0.07178. The effect of the Body Mass Index followed with an importance value of 0.07081 while the Histological grade was the next important variable with 0.06499. Further in this direction, the Histological Type followed with 0.06014 normalized importance levels whiles the CEA (0.04171), the Age (0.04074), Gender (0.03104) and Ascites (0.02813) followed in that order of importance.
**Figure 1: Overall Prediction Accuracy (ANN, CPH, PM)**

Figure 2 presents the prediction accuracy of the three models in term of estimating survivability among the patients. The ANN accurately forecasted 31% of patients to survive whiles 33% will not. On the other hand, the Cox Hazard Model accurately predicted 29% of the patients to survive whiles 31% will not. Finally, the PM market predicted 31% of the patients to survive whiles 31 percent will not survive. On the whole the prediction accuracy of the ANN was 64% whiles that of the CPH and the Prediction Market were 60% and 62% respectively. This implies that that whiles the ANN defeated the PM Model in predicting accuracy of survivability of patients with gastric cancer, it outperformed the Cox Hazard Model by 32 percentage points.

**LIST OF REFERENCES**


