

INTERNATIONAL JOURNAL OF SOFT COMPUTING



Career Path Decision Support System for Upper Basic School Students

I.O. Awoyelu, E.O. Oguntoyinbo and T.M. Awoyelu

Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria

Key words: Career path, decision support system, fuzzy logic, fuzzy K-nearest neighbor, undesirable

Abstract: Many students are faced with the challenges of deciding on suitable career path because decisions are characterized by a number of subjective judgements. Therefore, choosing a particular career path without first determining the suitability of the student as a fundamental step will yield an undesirable outcome. This paper aims at developing a career path decision making system for senior secondary schools. This model was developed based on the concept of fuzzy logic model. This model automatically converts crisp sets into fuzzy sets by using fuzzy k-nearest neighbour algorithm method. The model was implemented in the MATLAB environment. The performance of the model was evaluated using specificity and accuracy as performance metrics. The results obtained showed the accuracy of 90.22% and a specificity value of 96.97%. These results show that the system provides good support for decision making while eliminating the challenges of indecision and floundering that are characterized with choosing a career path among upper basic school students. The system will also serve as a tool in enhancing the work of career experts.

Corresponding Author:

I.O. Awoyelu

*Department of Computer Science and Engineering,
Obafemi Awolowo University, Ile-Ife, Nigeria*

Page No.: 58-69

Volume: 15, Issue 3, 2020

ISSN: 1816-9503

International Journal of Soft Computing

Copy Right: Medwell Publications

INTRODUCTION

Career is the sequence of occupations, jobs and positions held during the course of life^[1]. Choosing a career is one of the most significant decisions an individual is required to make^[2]. A career choice may satisfy an individual's needs, values and interests. The quality of an individual's life may also be influenced by the choice of career. It is also believed that the career decision always lead to important vocational outcomes in the future^[3].

All over the world, students often face the problem of decision making in career selection^[4-6]. Students who receive help to be able to explore career opportunities and planning programs of study, that relate to their career

interests are more likely to see how meaningful schools are^[7]. There are always insufficient counsellors to assist students in making appropriate vocational decisions based on their interest and self-concept^[8]. Abisoye *et al.*^[9] highlighted the problem of unpopularity and unawareness of the guidance and counselling department in secondary schools. The availability of a few number of human counsellors that attend to students in schools and the unavailability of counsellors in a good number of schools have further compounded the problem faced in the choice of career. As a result, many students have chosen their career path without receiving proper advice from suitable professionals and their decisions being influenced by peers and parent's preferences for certain careers.

Difficulties in the career decision-making process do lead to three major possible consequences. Firstly, it leads to the possibility of individual transferring the decision-making responsibility to others and refraining from making the decision themselves. Secondly, it leads to the failure to achieve their optimal career choice due to delay in decision-making. Thirdly, it leads to temporary unemployment^[10]. Students do not make informed career choices because they are largely influenced by their peers and parent's preference for certain careers^[11]. Other additional influencing factors are; students' perceptions about whether the subject is difficult or easy like or dislike for subject teachers and gender. Some parents tend to have a fixed idea of the vocation their children should choose irrespective of the child's intelligence, interest and other capabilities. Such wrong choices whether due to parental pressure or some other considerations have driven the naive youths into untold psychological problems; some of which have marred their future prospects^[12].

Decision on an appropriate career path for students transiting from upper basic school to senior secondary school is influenced by a number of judgements which are subjective. This leads to towing the wrong career path and/or under performance in a chosen career. In addition to the foregoing, there are always insufficient career counsellors. There is the need to correct this precarious situation, hence this paper aims at developing a career path decision support model for students. The specific objectives are to formulate a decision support model for career path decision making, implement the model and evaluate the model.

Literature review: Balogun *et al.*^[7] developed a knowledge assistant Decision Support System (DSS) for counselors to assist students in selecting the right discipline as they enter tertiary institutions, having considered the appropriate factors necessary for taking such decisions. The system was implemented on the platform of Visual Basic version 6.0 The DSS, tagged Career Master, consists of four databases namely; subject, study, pass and course. The subject database consists of all art, commercial and science subjects that are studied in secondary schools. The study database consists of all the study test questions used for the intelligent quotient assessment of the student. This also contains the answers to those questions. The Pass database consists of the password of the administrator form. The model management subsystem of the DSS is composed of the model base, model base management system, modeling language, model directory and model execution, integration and command processor. Model base contains routine and special statistical, counselling models, management, science and other quantitative models that

provide the analysis capabilities for the DSS. An appraisal of the implementation of this system was done by comparing it with the conventional method of carrying out career counselling. These suggestions are however based on some basic parameters used by human counsellors such as: Intelligence Quotient (IQ), hobbies, parent's and friends' influences and so on. Past academic records were not considered in determining suitability of students for a course in the university. IQ test was based on mathematics and literature test only. The system assisted students in selecting the right discipline as they enter tertiary institution. The language of implementation-VB is limited in terms of cross-platform compatibility. This implies that the DSS cannot be easily transferred and use in computer systems that use operating system other than Microsoft operating system.

Castellano and Martinez^[13] developed a decision support system that uses collaborative filtering for academic orientation technique. The system uses dataset of students from different Spanish secondary schools. The system offers advices and other additional information to facilitate their decision, so that, students will be successful in their academic choice. The system, based on the concept of collaborative recommender systems, supports academic orientation that analyzes the skills, attitudes and preferences of students. As a result, it computes relevant information that support user's decision concerning their academic future. It uses evaluative tools such as tests, essays, tasks, exercises and so on to evaluate students. The author's assumption is that the final result of the evaluative process will be a mark that reflects not only the student's knowledge but also their skills, preferences and tastes about their subject of choice. Factors such as past academic records were not considered.

Aslam and Khan^[14] also proposed a decision support system for guiding fresh students in selecting a faculty in Gomal University. The system uses a rule based technique to help students to select the best suitable faculty or major discipline based on their capabilities and abilities. The rule-based DSS works like an expert system which contains a general rule-base and an inference engine. The inference engine retrieves rules from the rule-base using C Language Integrated Production System (CLIPS) language. Rules for the DSS came from the knowledge of the human academic experts of Gomal University. The admission criteria of Gomal University, the past record of the students and their future result also form part of the rules used by the DSS. All the required criterias, abilities and capabilities for each faculty/major in Gomal University were collected and the knowledge was converted into facts and rules using CLIPS syntax and stored in the knowledge base of the DSS. The system was implemented using Microsoft VB because CLIPS

language does not support graphical interfaces. The system, designed using Microsoft Visual Basic was used for testing and measuring the student's capabilities like intelligence, understanding, comprehension, mathematical concepts, students past academic record, intelligence level and user's result. The result of the test is converted into text file which is opened by the CLIPS DSS for result interpretation. The use of the CLIPS allowed the forward chaining and rule based based on the Rete algorithm for pattern-matching^[15]. The recommendation was made based on student's background information, current academic record, intelligence test results and abilities module tests results. The DSS was able to identify the most suitable faculty or major course of study for the students based on the abilities and capabilities that were extracted from the result in the test module. The system was only useful for guiding fresh students in the selection of a faculty in the university. The system also gives a general test result for determining the best major for students without considering their past academic records.

Kostoglou etc., developed a web-based DSS that assists its users in being thoroughly informed about higher education studies and gets the users to choose their vocational prospects. The object-oriented approach with Unified Modeling Language (UML) was adopted to model the system. The system was implemented using Hypertext Preprocessor (PHP), MySQL and Javascript. Furthermore, Ajax technology was used to enhance the exchange of data with the database especially on-the-fly. To provide proper visualization, Google maps API was introduced into the system. Only job requirements and their relationship to higher educational studies were considered. Several factors such as student's academic records and interest that determines student suitability for a university course were not considered.

Oladokun and Oyewole^[16] worked on university-course admission choice using fuzzy logic. This study developed a decision support system for university admission seekers who are faced with the admission choice problem. Through literature search, interviews and expert knowledge, relevant factors that characterize the Nigerian University admission system were determined and the dynamics of their interaction were appropriately modeled. The fuzzy logic approach was adopted to linguistically and intuitively model the human reasoning and communication in the system. The equivalent fuzzy inference system of the decision process was developed. Model parameterization was carried out using information from the Nigerian University Admission System. The elements of the proposed fuzzy inference based decision system included a fuzzifier module, the fuzzy inference engine, the knowledge base and a defuzzification module. A two state variable model incorporating student ability

and interest was adopted. The fuzzy inference process that was used included the membership functions, fuzzy logical operations and fuzzy If-Then Rules. The rules set the basis for deriving the values of the output variable and choice viability, from the four input variables: Ability (A), Course Competiveness (CC), University Competiveness (UC) and Choice Preference (CP). The rules modeled how a career counseling expert will advise a candidate by comparing the candidate's academic ability with the perceived competition associated with his or her choice. The system cannot be used in determining career path for upper basic school students because the choice of career path had already been made. This was narrowed down to specific area of proficiency in terms of university course of choice. Past academic records were not also considered by the system.

Limitations of existing works: Most published research works on career decision making are to assist senior secondary school students in choosing the right discipline as they enter into tertiary institutions. In these systems, the choice of university courses had already been narrowed down to specific areas of proficiency based on the career path already chosen. This implies that a student that chooses art career path will be limited only to courses in the art field irrespective of whether such student is interested in science related fields. The choice of career path had been made. Therefore, these systems cannot assist students in upper basic school transiting to senior secondary school. There is also dearth of literature on career path decision support systems for students transiting from upper basic school to the senior secondary school.

MATERIALS AND METHODS

Proposed model: The dataset utilized in this study is Basic Education Certificate Examination (BECE) result, past academic records and personality profiles of students, which were collected from Basorun High School, Ibadan, Oyo State, Nigeria. It contains a total of (92) students made up of 58 girls and 34 boys. The dataset has (9) attributes based on personality profile, past academic record and the (7) core subjects that are taken at BECE. A sample of the dataset is as shown in Table 1. The dataset's output is a score on the four categories of student classes. This study attempts to identify the rules that distinguish the four different classes, that is, 'Science', 'Art', 'Commercial' and 'Repeat'. The method for assigning classes to students is dependent on supervised machine learning and model based diagnosis. The classification is based on the identification of patterns in the data. The task is to compare observations of the system modeled with that generated by the model. If the two deviate from each

Table 1: Sample dataset for basic education certificate examination result

Reg No.	Student name	EN	MT	BSCI	SS	CE	BT	BS
001	Abass Ezekiel	P	C	P	C	P	A	C
002	Abdulah WAR	P	C	P	C	P	C	C
003	Abiade Taiwo	C	C	C	C	P	C	C
005	Abideen Oduna	C	C	P	P	P	C	C
006	Abiodun Rofia	C	C	C	C	C	A	C
008	Abiola Benjam	P	P	C	C	C	A	C
016	Adeboye Samu	C	C	C	C	C	A	C
018	Adedokun BOS	C	P	C	P	P	C	P
020	Adegbola QOZ	C	C	C	P	C	A	C
023	Adekunle AYO	C	P	P	C	C	C	P
024	Adelalu Tosin	P	P	P	C	C	C	P
029	Adeoye Rofia	C	P	C	C	C	C	C
033	Adesina Dami	C	P	P	C	C	C	P
034	Adesina Suliat	C	C	C	C	C	C	P
036	Adeyemi Oreol	P	P	P	P	P	C	P
038	Adeyemo Emm	P	P	C	P	C	C	C
039	Adeyemo Debo	P	C	C	C	C	A	P
045	Afolabi Esthe	P	P	P	P	P	C	P
046	Afolalu Titila	P	C	C	C	C	A	C
050	Ajibade Taiwo	P	P	C	C	C	C	P
052	Ajiboye Tabith	C	C	C	C	C	A	C
054	Akanbi Oyinad	P	P	C	P	C	C	C
056	Akande Bukol	C	P	C	C	C	C	C
061	Akingbala OPE	P	P	C	C	C	A	C

Table 2: Training and testing parameters for the fuzzy-logic-based classifier

Classes (set)	No. of instances	No. of features	No. of classes
Training	46	9	4
Test	46	9	4

Table 3: Different categories of data and descriptions

Data category (Classes)	Description
1	Science
2	Art
3	Commercial
4	Repeat

Table 4: Fuzzy set for BECE

Linguistic variables	Notation	Range	Crisp values
Distinction	A	70-100	1
Credit	C	50-59	2
Pass	P	40-49	3
Fail	F	0-39	4

other an anomaly is considered detected. Fuzzy KNN algorithm was employed as the classifier to assist in deciding the appropriate career path for each student based on the three identified influencing factors. The training and testing parameters and conditions are defined in Table 2. Data was segmented into 4 distinct classes, Classes 1 to 4 were assigned for ‘Science’, ‘Arts’, ‘Commercial’ and ‘Repeat’ and defined in Table 3. The model estimation was carried out in the MATLAB environment by comparing the true labels with the classifier assigned crisp output values. Firstly, the data was randomly stratified to avoid the data being biased and afterwards, the FKNN-based model is firstly trained on the training sets to tune the parameters via. stratified 2-fold cross validation in order to obtain the optimal

parameter pair (k, m) and then the resulting optimal FKNN model was employed in carrying out the classification tasks.

Development of the fuzzy-based career path decision support system: Mamdani fuzzy inference system was used to develop the fuzzy rule-based approach. It is used to obtain the values for the output variable from the input variables provided to the fuzzy inference system structure. The system architecture of the fuzzy based career path decision support system is presented in Fig. 1. The figure shows the Basic Education Certificate Examination (BECE) result combined with past academic record and personality profile as the input variables to give appropriate career path options as the output variable. It consists of four operating mechanisms namely: fuzzification, fuzzy knowledge base, inference engine and defuzzification.

Fuzzification: The fuzzification of variables involved the conversion of crisp values that was assigned to the input variables and into their respective fuzzy logic linguistic variables. These were formally specified as a mathematical model called a membership function. By fuzzification, the crisp value is converted into fuzzy input value with triangular membership function. Triangular membership function was used for converting the crisp set into fuzzy set due to their simple formula and computational efficiency. Table 4 is the fuzzy set for the input variables-BECE showing linguistic variables which are distinction, credit, pass and fail, its notations, ranges and crisp values. These linguistic variables

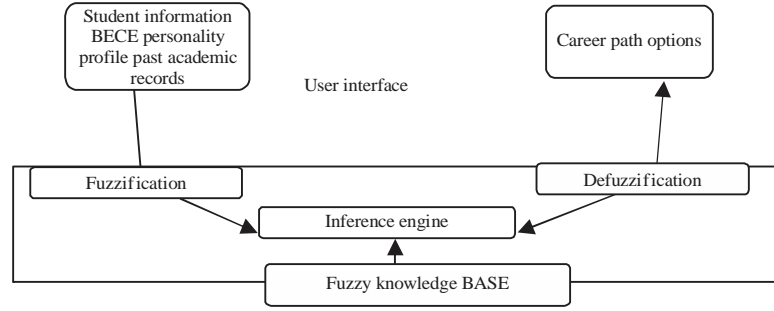


Fig. 1: System architecture for career path decision support system

Table 5: Fuzzy set for past academic records

Linguistic variables	Notation	Range	Crisp values
Excellent	A	70-100	1
Very good	B	60-69	2
Good	C	50-59	2
Fair D	45-49	3	
Poor E	40-44	3	
Fail F	0-39	4	

represent the gradual transition from high to low. They are words from natural language instead of numerical values.

Table 5 is the fuzzy set for the past academic records, showing linguistic variables which are excellent, very good, good, fair, poor and fail, its notations and crisp values. These linguistic variables represent the gradual transition from high to low. Table 6 is the fuzzy set for the input variables, students' personality profile showing the personality type with its corresponding career path and the linguistic variable such as like very much, like, indifferent, dislike, dislike very much. Fuzzy set allows its members to have different grades of membership in the interval [0, 1].

The K-Nearest Neighbor (KNN) algorithm is a supervised learning algorithm or a simple instance-based classifier which carries out its classification tasks by assigning the class or output of a new sample according to the most common class amongst its k nearest neighbors. The classification is based on majority of k-nearest neighbor category. In a given sample, a K number of objects nearest to the samples point are decided. The classification is carried out using majority vote among the classification of the k objects.

As a better variant of the KNN algorithm, the Fuzzy K-Nearest Neighbour (FKNN) classifier designed is a classification algorithm that overcomes the limitations of the KNN classifier by incorporating the fuzzy set theory into KNN. In FKNN, rather than assigning individual classes as in the original KNN, the fuzzy memberships of samples are assigned to different classes. The class assigned to be the winner is the one which has the highest membership degree. For pattern x, the FKNN algorithm attributes to a membership vector as a function of the pattern's distance from its k-nearest neighbors. This

guarantees that no arbitrary assignments are made. The first step to be carried out for the FKNN algorithm is computing the fuzzy partition matrix $U = (u_{ij})$ from the memory where a set of n training sample vectors (x_1, \dots, x_n) are stored. j is represented as the vector index $(j = 1, 2, \dots, n)$ where n is the number of training samples, the variable i represents the class index $(i = 1, 2, \dots, C)$ and C is the number of classes. For each training case x, we identify its k nearest neighbors by computing Euclidean distances. The membership degree of the sample vector x_j in the class i is as given in Eq. 1 where n_i is the number of neighbors discovered to belong to the class i and $c(x_j)$ represents the class label of the sample vector x_j :

$$u_{ij}(x) = u_i(x_j) = \begin{cases} 0.51 + (n_i / k) \times 0.49, & \text{if } c(x_j) = i \\ (n_i / k) \times 0.49, & \text{if } c(x_j) \neq i \end{cases} \quad (1)$$

n_i represents labeled reference patterns amongst the k closest labeled reference patterns which are labeled in class x_i and j ranges from 1 to n. The memberships are shared among the classes. It can be clearly seen that u_{ij} is an element of the C-by-n matrix U. Also, worthy of note is the fact that the purpose of Eq. 1 is to allocate higher fuzzy membership values to the training samples that stay away from the decision boundary and lower fuzzy memberships values to the patterns which lie in the neighbourhood of the decision boundary. It is because the information supplied by the samples in the region close to the decision surface is more uncertain than those provided by other samples. Since, u_{ij} is a fuzzy membership grade or value of the sample x_j in the class i, u_{ij} must satisfy the following properties:

$$u_{ij} \in [0,1] \quad (2)$$

$$\sum_{i=1}^c u_{ij} = 1 \quad (3)$$

Where:

$$0 < \sum_{j=1}^n u_{ij} < n$$

Table 6: Fuzzy set for students' personality profile

Questions (section A-J)	Personality types	Career path	Linguistic variable
1	Outdoor	Science	Like very much, like, indifferent, dislike, dislike, dislike very much
2	Mechanical	Science	Like very much, like, indifferent, dislike, dislike, dislike very much
3	Computational	Science	Like very much, like, indifferent, dislike, dislike, dislike very much
4	Scientific	Science	Like very much, like, indifferent, dislike, dislike, dislike very much
5	Persuasive	Art	Like very much, like, indifferent, dislike, dislike, dislike very much
6	Artistic	Art	Like very much, like, indifferent, dislike, dislike, dislike very much
7	Literary	Art	Like very much, like, indifferent, dislike, dislike, dislike very much
8	Musical	Art	Like very much, like, indifferent, dislike, dislike, dislike very much
9	Social Service	Commercial	Like very much, like, indifferent, dislike, dislike, dislike very much
10	Clerical	Commercial	Like very much, like, indifferent, dislike, dislike, dislike very much

The FKNN algorithm is to assign fuzzy memberships of the unknown sample x to different classes according to the following equation:

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} (1 / \|x - x_j\|^{2(m-1)})}{\sum_{j=1}^k (1 / \|x - x_j\|^{2(m-1)})} \quad (4)$$

$$C(x) = \arg_{i=1}^c \max(u_i(x)) \quad (5)$$

where, $i = 1, 2, \dots, C$ and $j = 1, 2, \dots, k$. In this context, j represents the j th sample vector among the k nearest neighbors of x . C is the number of classes; k denotes the neighboring size. The fuzzy strength parameter m is used to ascertain how heavily the distance is weighted when calculating each neighbor's contribution to the membership value and is determined a-priori from the training data. The value of m is usually taken as $m \in (1, \infty)$. $\|x - x_j\|$ denotes the distance between x and its j th nearest neighbor x_j . In this write up, Euclidean metric is used as the distance measurement. And u_{ij} which denotes the membership degree of the sample vector x_j in the class i is computed in the first step of the algorithm. There are two types of labeling namely, fuzzy labeling, otherwise known as soft labeling and crisp labeling. This study is based on the fuzzy labeling methods where is the membership of the test vector x , to class i , $\|x - x_j\|$ is the Euclidean distance between the test vector x and the k -th nearest neighbor vector x_j and m is a real number greater than or equal to 1.0 that sets the "strength" of the fuzzy distance function. The FKNN is a feature-based classifier system. In this study, features were extracted from the segmented training dataset (50% for training and 50% for testing) and fed into the classifier which assigns a user-defined class label to the obtained results. The FKNN classifier system possesses the capability to adjust itself to the ambiguity in the training and test data. The classifier ascribes a membership value to the unlabeled object and this provides the system with information which is useful in estimating the accuracy of the decision. This fuzzy membership value identifies what fraction of an unlabeled object resides in each of the defined classes. The FKNN algorithm is as shown in Algorithm 1.

Optimization of the classifier algorithm is carried by looping through the process of selecting a locally optimal feature selection vector, k and fuzzy strength parameter, m for each class pair, ij . The process begins by ranking the features that correspond to the training data input by the user and by specifying a constant initial fuzzy strength value for all class labels. The ranking process is deterministic for the feature set and is not repeated as the algorithm iterates through the training process. After feature-ranking an estimate of the classifier's stratified 2-fold cross validated performance is generated for the ranked features beginning with the best feature, the best two features, best three features, etc., until a list of performance versus weight vectors is derived for every pair-wise class comparison. From this list, the best weight vector, k is determined for each class pair, ij . Once the best features have been selected, the stratified 2-fold cross validated performance is generated again as the system increments the fuzzy strength parameter, m for each class pair, ij . The optimal values for m are then chosen to represent each class pair. This completes the first iteration through the training procedure. A trained classifier comprises of the training vectors, the optimal weight vectors, and the optimal fuzzy strength parameters. If new vectors are later added to the training set, the weight vector and fuzzy strength should be updated.

Algorithm1; The FKNN Algorithm:

Input: The training set X with the labeled patterns $\{x_j, j=1, 2, \dots, n\}$ and the test pattern y

Output: Class label of y and confidence for each class label

1. For $i = 1, 2, \dots, n$
2. Compute the distance from x_i to y using the Euclidean distance
3. If i to k
4. Include x_i in the set of k nearest neighbors
5. Else If (x_i is closer to y than any previous nearest neighbors)
6. Discard the farthestmost of the k nearest neighbors
7. Include x_i in the set of k nearest neighbors
8. End If
9. End for
10. For $c = 1-C$
11. Compute $u_i(x)$ using Eq. 4
12. End for
13. Crisp class label of y is assigned to the class with which it has the highest membership value according to Eq. 5

Fuzzy knowledge base: The rule base and the database are jointly referred to as fuzzy knowledge base. Fuzzy rules were developed based on the knowledge elicited from the career experts. The rule base contains a collection of rules which represent the knowledge about the application domain specified by the domain expert. The career counselor acts as the domain expert by providing data about personality profile and academic records and the rules required to determine appropriate career path. The <condition> and <action> parts of each rule are denoted as the antecedent and consequent parts, respectively. The antecedent part can be a simple or complex logical combination of conditions and more than one action may be specified in the consequent part. The rules are constructed and set the basis for deriving the values of the output variables. It models how a career expert will advise a student by comparing the basic education certificate examination result, past academic records and personality profile. The database component contains an ensemble of membership functions of the fuzzy sets used in the fuzzy rules. The fuzzy rules are the following:

IF (((BECE ((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((bsci,'A')|(bsci,'C'))&((ss,'A')|(ss,'C'))&((bt,'A')|(bt,'C'))'C'))&((bs,'A')|(bs,'C'))|(pastacademicrecords(((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((bsci,'A')|(bsci,'C'))&((ss,'A')|(ss,'C'))&((bt,'A')|(bt,'C'))&((ce,'A')|(ce,'C')) & ((bs,'A')|(bs,'C')))) THEN careerpath is 'Science-Art-Commercial' ELSE IF (((BECE ((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((ss,'A')|(ss,'C'))&((ce,'A')|(ce,'C'))&((bs,'A')|(bs,'C'))|(past academic records((en,'A')|(en,'C'))& ((mth,'A')|(mth,'C'))&((ss,'A')|(ss,'C'))&((ce,'A')|(ce,'C'))&((bs,'A')|(bs,'C')))) THEN careerpath is 'Art-Commercial'

ELSE IF (((BECE ((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((ce,'A')|(ce,'C')) & ((bs,'A')|(bs,'C'))|(past academic records ((BECE ((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((ce,'A')|(ce,'C')) & ((bs,'A')|(bs,'C')) THEN careerpath is 'Commercial' ELSE IF (((BECE ((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((ss,'A')|(ss,'C'))&((ce,'A')|(ce,'C'))&((pastacademicrecords ((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((ss,'A')|(ss,'C'))&((ce,'A')|(ce,'C')) THEN careerpath is 'Art' ELSE IF (((BECE((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((bsci,'A')|(bsci,'C'))&((bt,'A')|(bt,'C'))|(past academicrecords ((en,'A')|(en,'C'))& ((mth,'A')|(mth,'C'))&((bsci,'A')|(bsci,'C'))&((bt,'A')|(bt,'C')) THEN careerpath is 'Science' ELSE IF (((BECE ((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((bsci,'A')|(bsci,'C'))&((ss,'A')|(ss,'C'))&((bt,'A')|(bt,'C'))&((ce,'A')|(ce,'C'))|(pastacademicrecords (((en,'A')|(en,'C'))& ((mth,'A')|(mth,'C'))&((bsci,'A')|(bsci,'C'))&((ss,'A')|(ss,'C'))&((bt,'A')|(bt,'C'))&((ce,'A')|(ce,'C')) THEN careerpath is 'Science-Art' ELSE IF (((BECE ((en,'A')|(en,'C'))&

((mth,'A')|(mth,'C'))&((bsci,'A')|(bsci,'C'))&((bt,'A')|(bt,'C'))&((ce,'A')|(ce,'C'))|(pastacademicrecords (((en,'A')|(en,'C'))&((mth,'A')|(mth,'C'))&((bsci,'A')|(bsci,'C'))&((ss,'A')|(ss,'C'))&((bt,'A')|(bt,'C'))&((ce,'A')|(ce,'C')) THEN careerpath is 'Science-Commercial' ELSE IF (((BECE ((en,'F')&((mth,'F') THEN 'repeat class' End.

Legend: en-English; mth-Mathematics; bsci-Basic Science; ss-Social Studies; ce-Civiic Education; bs-Business Studies; bt-Basic Technology.

Personality profile rules

```
if (pt, 'outdoor')
    pp = 'Science'
else if (pt, 'mechanical')
    pp = 'Science'
else if (pt, 'computational')
    pp = 'Science'
else if (pt, 'scientific')
    pp = 'Science'
else if (pt, 'persuasive')
    pp = 'Art'
else if (pt, 'Artistic')
    pp = 'Art'
else if (pt, 'literary')
    pp = 'Art'
else if (pt, 'musical')
    pp = 'Art'
else if (pt, 'social')
    pp = 'Commercial'
else if (pt, 'clerical')
    pp = 'Commercial'
end
a. If (Career_Path_Options is science or art or commercial class) and (if personality_profile is science) then (Career_Path_Options is science class)
b. If (Career_Path_Options is art or commercial class) and (if personality_profile is commercial) then (Career_Path_Options is commercial class)
c. If (Career_Path_Options is commercial class) and (if personality_profile is commercial) then (Career_Path_Options is commercial class)
d. If (Career_Path_Options is art class) and (if personality_profile is art) then (Career_Path_Options is art class)
e. If (Career_Path_Options is science class) and (if personality_profile is science) then (Career_Path_Options is science)
f. If (Career_Path_Options is science or art class) and (if personality_profile is science) then (Career_Path_Options is science)
g. If (Career_Path_Options is science or commercial class) and (if personality_profile is science) then (Career_Path_Options is science class)
h. If BECE (English_language is P) and (mathematics is P) then repeat class
```

Inference engine: This has the rule based decision support capacity and information processing component of the career path decision support system for schools. It scans through the fuzzy knowledge base to evaluate the different rules in the knowledge base. Fuzzy inference

(reasoning) is the actual process of mapping from a given input to an output using fuzzy logic. A fuzzy response is generated from these fuzzified inputs using a rule based inference system. The inference engine is the mechanism that performs the reasoning and searching in the fuzzy knowledge base. It implements the fuzzy reasoning by combining the fuzzified inputs with the rules to obtain a fuzzy output. When the inference process is complete, the resulting data for each output of the fuzzy classification system is a collection of a single, aggregate fuzzy set.

Defuzzification: When the inferencing is over, there is need to compute a single value to represent the outcome. This process is called defuzzification. This involves the conversion of the fuzzy linguistic values to the original crisp values of the output variable. The implication result obtained for each rule are aggregated and defuzzified to obtain single crisp values. In this study, the defuzzification method used is the centre of area, otherwise known as Centroid of Area (COA) defuzzification technique. Centroid of area zCOA is defined in Eq. 6:

$$z_{COA} = \frac{\int_z \mu_A(z)z dz}{\int_z \mu_A(z)z dz} \quad (6)$$

where, $\mu_A(z)$ is the aggregated output MF. This is the most widely adopted defuzzification strategy which is reminiscent of the calculation of expected values of probability distributions.

Classification phase using FKNN: The model implemented its classification tasks by making use of the randomly stratified data. It contains two main sub-modules. In FKNN classification, the first step is to set up all the model parameters. Because the fuzzy strength parameter has a major impact on the performance of the FKNN-based model, an experimental outline was drawn in order to select the optimal fuzzy strength parameter for the FKNN classifier. An interval of (1, 2) was chosen and continually increased with steps of 0.01 for the fuzzy strength parameter m and subsequently classification performance was confirmed by way of the 2-fold CV analysis on various values of nearest neighbors k . For every value of m in the interval (1, 2), the average accuracy obtained by FKNN by the way of Cross Validation (CV) analysis, and eventually, the one with the highest average accuracy was selected as the optimal fuzzy strength parameter. After choosing the optimal fuzzy strength parameter, the FKNN classifier was employed to compute the classification accuracy and then an average of the results was taken. The pseudo-code for the classification phase is given in Algorithm 2.

Algorithm 2; The pseudo-code for the classification phase:

```

/*performance estimation by using n-fold CV where n = 2*/
1. Begin
2. For i = 1:Mmax
3. For j = 1:k
4. Training set = k - 1 subsets
5. Test set = remaining subset
6. Train the FKNN model on the training set to find the optimal fuzzy strength parameter m when the neighborhood size k is set from 1 through 15
7. Test it on the test set and assigns the accuracy to V(j), where V is a vector whose element is the corresponding accuracy obtained by each folder
8. End for
9. Compute the mean value of vector V, and store the mean CV accuracy in the vector M(i)
10. End for
11. Get the optimal m value whose corresponding mean CV accuracy is the highest in M(i)
12. End
13. Begin
14. For l = 1:k
15. Training set = k - 1 subsets
16. Test set = remaining subset
17. Train the FKNN model on the training set using the obtained optimal parameter combination
18. Test it on the test set and save the mean CV accuracy
19. End for
20. Return the average classification accuracy rates of FKNN over 1 test set
21. End
    
```

The fuzzy inference system for career path decision support system is shown in Fig. 2. It involves the fuzzification of all input variables which are personality profile, past academic records and each compulsory seven subjects in BECE. All the input variables were fuzzified using triangular membership function because of its simplicity and clarity. It also involves the development of the rule base for the inference engine, aggregation of the results processed by the inference engine for the inputs value presented to the fuzzy inference system and defuzzification of the output using triangular membership function.

Figure 3 shows the membership function of past academic records inputs showing the four different classes for the different career path options. The membership function is a curve that allows a graphical representation of a fuzzy set. It defines how each point in the input space is mapped to a membership value between 0 and 1. A membership function is used to quantify a linguistic terms.

Figure 4 and 5 show the triangular membership function of BECE subjects-Mathematics and Basic Science inputs and the corresponding range for each notation, F (Fail), C (Credit), P (Pass) and A (Distinction). Figure 6 shows the output membership function for the career path. The rule base developed for the inference engine is shown in Fig. 7. These rules obtained from career experts are expressed as linguistic rules. The rule

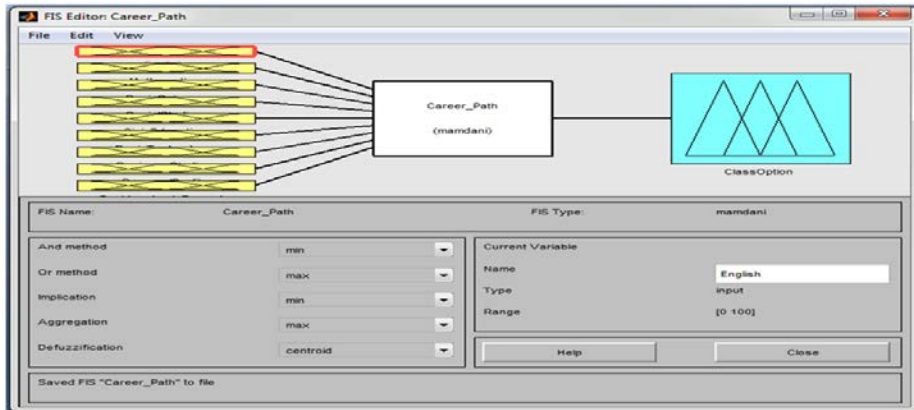


Fig. 2: Fuzzy inference system for career path decision support system

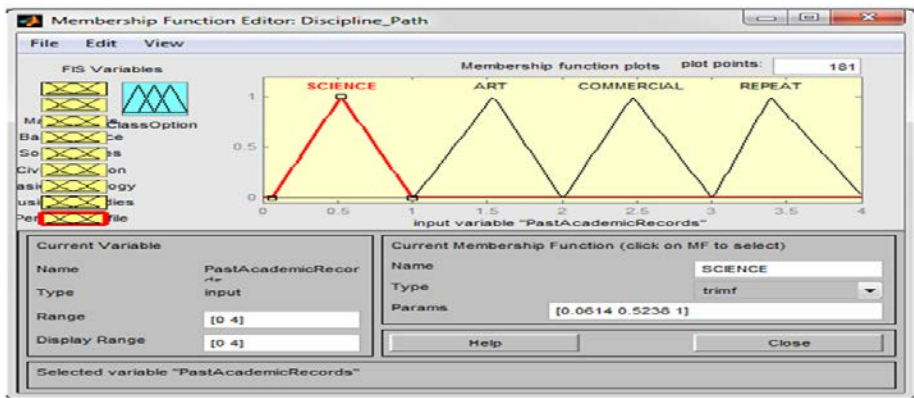


Fig. 3: Membership function for past academic records

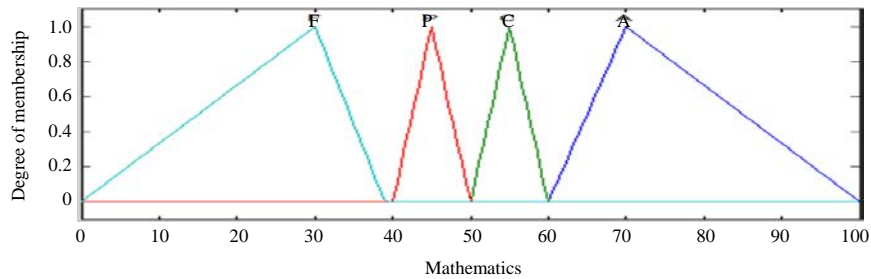


Fig. 4: Membership function for BECE (Mathematics)

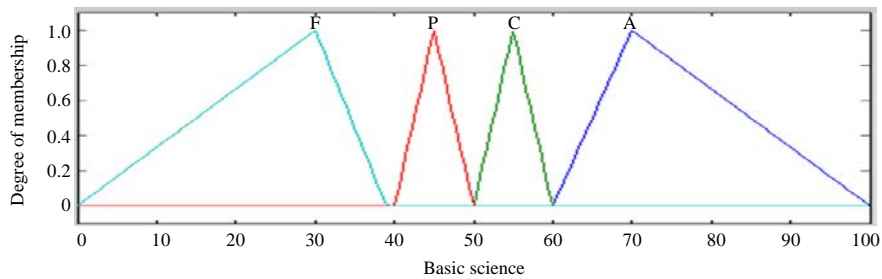


Fig. 5: Membership function for BECE (Basic Science)

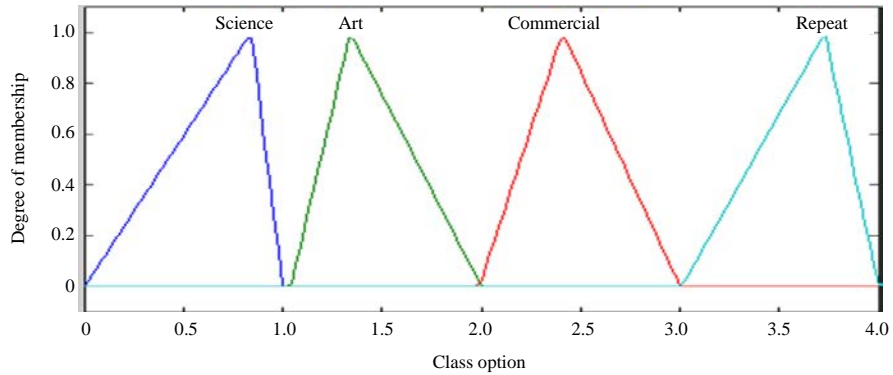


Fig. 6: Output membership function for career path

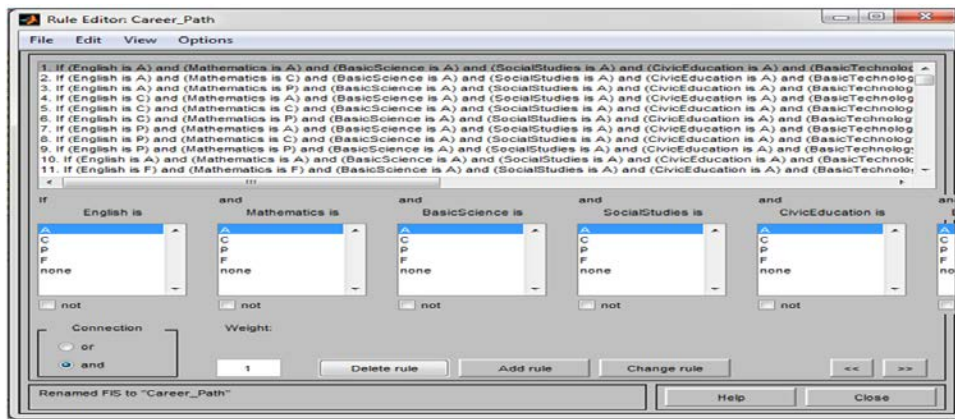


Fig. 7: Snapshot of some rules developed for the inference engine

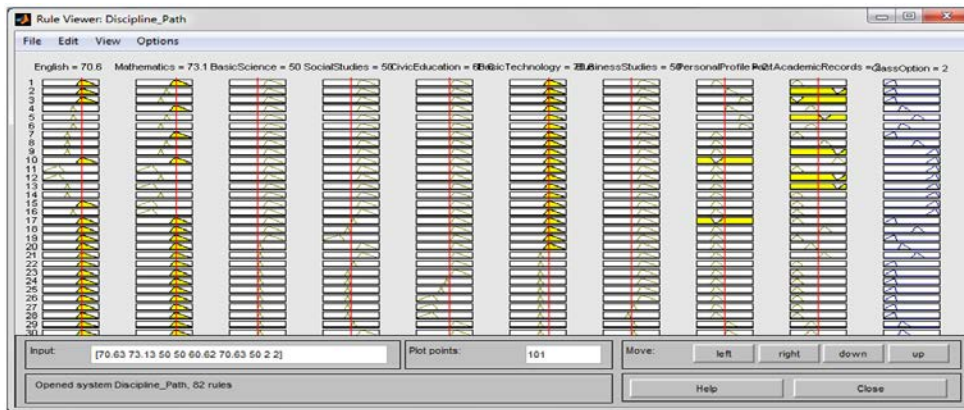


Fig. 8: Rule viewer showing the inputs and defuzzified output

viewer of the career path decision support system is shown in Fig. 8, showing the inputs and defuzzified outputs. The input values can be changed by clicking and dragging the input vertical lines to give defuzzified outputs.

RESULTS AND DISCUSSION

System evaluation: In order to properly evaluate the performance of the fuzzy inference system classifier model, a stratified 2-fold cross validation model was used

in which the folds are selected, so that, each fold contains roughly the same proportions of class labels. The input data was partitioned into two equal datasets. Accuracy and specificity were used to test the performance of the fuzzy based career path decision support system. Accuracy is calculated as the number of all correct predictions divided by the total number of the dataset. Specificity is calculated as the number of correct negative predictions divided by the total number of negatives. The performance of the model was evaluated using the data in the confusion matrix. It reports the number of True Positives (TP), False Negatives (FN), True Negative (TN) and False Positives (FP) for a four class classifier. Each row in the confusion matrix represents an actual class and each column represents a predicted class. TP is the number of students whose career path was correctly predicted; FN is the number of students whose career path was incorrectly predicted; TN is the number of students whose career path was correctly predicted and FP is the number of students whose career path was incorrectly predicted. The predicted accuracy and specificity are calculated as follows:

$$\text{Predicted accuracy} = ((TP+TN)/(TP+FN+FP+TN))*100\% \quad (7)$$

$$\text{Specificity} = (TN/(FP+TN))*100\% \quad (8)$$

Performance evaluation was carried out using two performance evaluation metrics. The methods used to validate the system performance are accuracy (also known as correct classification rates) and specificity. Accuracy measures the ratio of correct predictions to the total number of cases evaluated. Specificity (also called the true negative rate) measures the proportion of negatives that are correctly identified. Accuracy measures the probability that the system can correctly decides appropriate career path for upper basic school students. The total number of test cases is (46). As displayed in Fig. 9 and 10:

- Accuracy for fold 1 = $((TP+TN)/(TP+FN+FP+TN))= 91.30\%$
- Accuracy for fold 2 = $((TP+TN)/(TP+FN+FP+TN))= 89.13\%$
- Average Accuracy = 90.22%
- Specificity = $(TN/(FP+TN)*100\%) = 96.97\%$

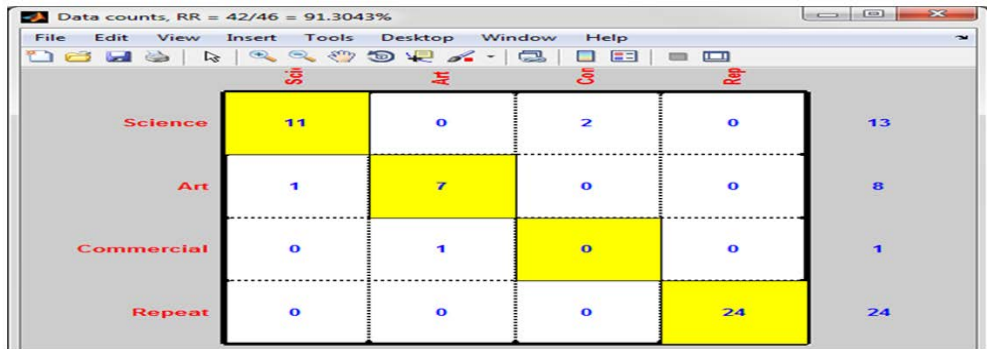


Fig. 9: Confusion matrix displaying the crisp outputs for fold 1

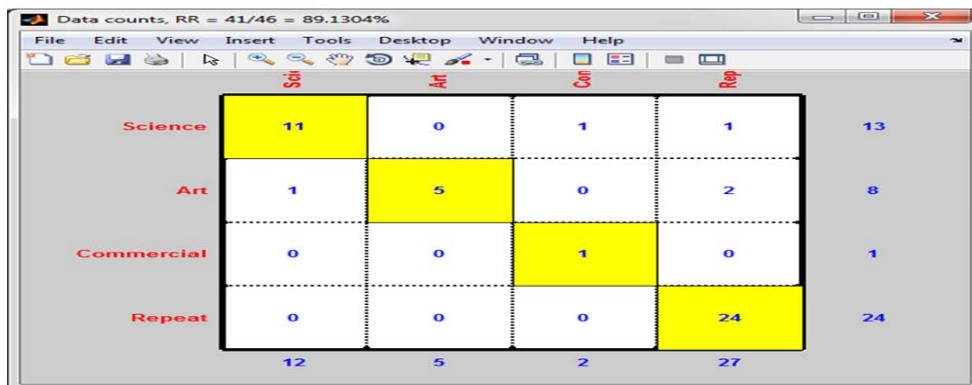


Fig. 10: Confusion matrix displaying the crisp outputs for fold 2

The performance evaluation of the model shows 90.22% accuracy in deciding appropriate career path for students.

CONCLUSION

Many students are faced with the challenges of deciding on suitable career path because decisions are characterized by a number of subjective judgements. Therefore, choosing a particular career path without first determining the suitability of the student as a fundamental step will yield an undesirable outcome. This research work addressed this issue by developing a career path decision making system. The resulting system will also serve as a tool in enhancing the work of career experts.

REFERENCES

01. Ezeani, N., 2013. Career choice: A basic issue in primary and secondary school level. *Arabian J. Bus. Manage. Rev.*, 1: 18-28.
02. Gati, I. and B. Asher, 2001. Differential diagnosis and treatment of career indecision. *Personal Guidance J.*, 62: 27-29.
03. Walker, T.L. and T.J. Tracey, 2012. The role of future time perspective in career decision-making. *J. Voc. Behav.*, 81: 150-158.
04. Bandura, A., C. Barbaranelli, G.V. Caprara and C. Pastorelli, 2001. Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Dev.*, 72: 187-206.
05. Issa, A.O. and K.I.N. Nwalo, 2008. Factors affecting the career choice of undergraduates in Nigerian library and information science schools. *Afr. J. Lib. Arch. Inf. Sci.*, 18: 27-37.
06. Razak, T.R., M.A. Hashim, N.M. Noor, I.H. Abd Halim and N.F.F. Shamsul, 2014. Career path recommendation system for UiTM Perlis students using fuzzy logic. *Proceedings of the 2014 5th International Conference on Intelligent and Advanced Systems (ICIAS), June 3-5, 2014, IEEE, Kuala Lumpur, Malaysia*, pp: 1-5.
07. Balogun, V.F., A.F. Thompson and M. Tech, 2009. Career master: A Decision Support System (DSS) for guidance and counseling in Nigeria. *Pacific J. Sci. Technol.*, 1: 337-354.
08. Otta, F.E. and N.O. Williams, 2012. Self concept and vocational interest among secondary school students (Adolescents). *Asian J. Social Sci. Hum.*, 1: 38-48.
09. Abisoye, O., I. Alabi, O. Ganiyu, O. Abisoye and J. Omokore, 2015. A web-based career guidance information system for pre-tertiary institution students in Nigeria. *Int. J. Sci. Res. Sci. Eng. Technol.*, 1: 229-240.
10. Gati, I., N. Saka and M. Krausz, 2001. Should I use a computer-assisted career guidance system? It depends on where your career decision-making difficulties lie. *Br. J. Guidance Counselling*, 29: 301-321.
11. Lazarus, K.U. and C. Ihuoma, 2011. The role of guidance counsellors in the career development of adolescents and young adults with special needs. *Br. J. Arts Social Sci.*, 2: 51-62.
12. Kinanee, J.B., 2009. Factors in the career decision-making of nurses in Rivers State of Nigeria: Implications for counselling. *Int. J. Psychol. Counselling*, 1: 134-138.
13. Castellano, E.J. and L. Martinez, 2009. A web-decision support system based on collaborative filtering for academic orientation. Case study of the Spanish secondary school. *J. UCS.*, 15: 2786-2807.
14. Aslam, M.Z. and A.R. Khan, 2011. A proposed decision support system/expert system for guiding fresh students in selecting a faculty in Gomal university, Pakistan. *Ind. Eng. Lett.*, 1: 33-41.
15. Zhengxin, C., 2000. *Computational Intelligence for Decision Support*. CRC Press, Florida, USA.,
16. Oladokun, V.O. and D.I. Oyewole, 2015. A fuzzy inference based decision support system for solving the university-course admission choice problem. *Int. J. Comput. Appl.*, 112: 1-7.