Analyse of 3D geometrical STEP file for feature recognition

MSc Research Project

MSc in AI for Business

Adil Abdullah

Student ID: x23114631

School of Computing

National College of Ireland

Supervisor: Brian Byrne

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……………Adil Abdullah………………………………………………………………

Student ID: ……………x23114631………………………………………………………………………

Programme: ……MSc in AI for Business …………….. **Year:** ……2023-24……..

Module: ……………………Practicum 2……………………………………………………………

Supervisor: ……………Brian Byrne……………………………………………………………………….

Submission Due Date: ……………16-09-2024……………………………………………………………………

Project Title: …. Analyse of 3D geometrical STEP file for feature recognition…

Word Count: ………6471… **Page Count:** ………………22……………….……….

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………………ADIL ABDULLAH…………………………………………

Date: …………………16-09-2024…………………………………………..

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Analyse of 3D geometrical STEP file for feature recognition

Adil Abdullah
X23114631

Abstract

The areas of manufacturing, engineering and computer-aided design (CAD) help determine accurate specifications based on 3D geometric data. This study presents a deep analysis of STEP files of 3D geometry with an emphasis on feature recognition. Complex 3D models are represented in STEP files, which is a serious problem. These data structures and geometric objects can be difficult to code due to their complexity and variety. Overview of CNN and other machine learning algorithms. STEP data uses traditional feature-based methods to identify and classify holes, grooves and pockets. In this case, features such as holes, grooves and pockets are identified and classified from the STEP data using traditional feature-based methods. CNN excels at recognizing features from voxel or mesh representations for analysis. Simple geometric shapes remain competitive in terms of computational efficiency and accuracy even when treated with traditional methods. This study highlights the trade-off between accuracy, computational requirements and model interpretation.

Keywords: STEP File, 3D Geometrical Model, Feature recognition in intelligence, 3D CAD Modelling

1. Introduction:

3D CAD models, especially STEP files, are processed to identify features for computer-aided design and manufacturing. This analysis automatically identifies and the 3D model makes its geometric features explicit, making the design process more efficient. [1]The ISO standard for conversion of geometric data and 3D products, known as STEP ISO 10303, is widely used in many CAD systems. The shape, texture, and properties of 3D objects are described in detail in STEP files. The 3D geometric data is processed to identify the features it contains. This is an important property in many industrial and engineering applications. 3D model features are identified and extracted, including holes, grooves, voids, protrusions, and other geometric objects. Many subsequent processes depend on this property.

Header Section:

```
ISO-10303-21;
HEADER;
FILE_DESCRIPTION (( 'STEP AP203' ),
    '1' );
FILE_NAME ('mac-003-006_Halter.STEP',
    '2012-11-28T08:14:58',
    ( 'admin' ),
    ( 'Hewlett-Packard Company' ),
    'SwSTEP 2.0',
    'SolidWorks 2010',
    '' );
FILE_SCHEMA (( 'CONFIG_CONTROL_DESIGN' ));
ENDSEC;
```

Data Section:

```
DATA;
#1 = CARTESIAN_POINT ( 'NONE',  ( 26.50000000000000400, -27.75000000000000400, 55.00000000000000000 ) ) ;
#2 = PRODUCT ( 'mac-003-006_Halter', 'mac-003-006_Halter', '', ( #1148 ) ) ;
#3 = PERSON ( 'UNSPECIFIED', 'UNSPECIFIED', 'UNSPECIFIED', ('UNSPECIFIED'), ('UNSPECIFIED'), ('UNSPECIFIED') ) ;
#4 = CIRCLE ( 'NONE', #1399, 17.50000000000000400 ) ;
#5 = EDGE_LOOP ( 'NONE', ( #1820, #1815 ) ) ;
#6 = LINE ( 'NONE', #1235, #259 ) ;
#7 = CIRCLE ( 'NONE', #1403, 2.749999999999992000 ) ;
#8 = LINE ( 'NONE', #1221, #10 ) ;
#9 = LINE ( 'NONE', #349, #13 ) ;
#10 = VECTOR ( 'NONE', #1227, 1000.000000000000000 ) ;
#11 = VECTOR ( 'NONE', #801, 1000.000000000000000 ) ;
#12 = VECTOR ( 'NONE', #350, 999.9999999999998900 ) ;
#13 = VECTOR ( 'NONE', #352, 1000.000000000000000 ) ;
#14 = LINE ( 'NONE', #268, #78 ) ;
#15 = LINE ( 'NONE', #798, #16 ) ;
```

Fig 1: STEP file data sections

The industry is moving towards automation and intelligence as part of the industrial paradigm. [2]The importance of efficiently and accurately identifying features in 3D models has increased significantly. Advanced manufacturing technology is integrated into CAD models through automatic feature recognition, ensuring a seamless transition from design to production. STEP 3D geometry files are analysed to identify critical application features that bridge the gap between digital design and physical manufacturing. [3]Engineers and manufacturers can identify the characteristics of these models, allowing them to optimize the production process and improve quality control and product design. These advanced computational algorithms and methods are constantly improving and pushing the boundaries of what is possible in this field. Functional identification has become an essential tool in modern engineering and manufacturing processes.

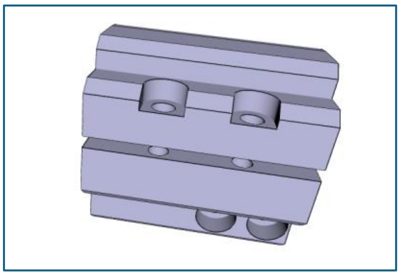

Fig 2: View of STEP file model

2. Literature Review:

D. Sreeramulu1*, C.S.P.Rao2 [1] In this study the purpose is to extract the functions of the rotating part, regardless of the symmetry of the rotating part. Here is a method for determining curvature properties. Identify rotating objects, analyse their properties, determine their significance in your data, apply rotating object detection algorithms, and validate the results. Here, we use several algorithms related to machine learning for recognize segment features. A geometry extraction algorithm was developed to extract geometric and information from STEP file regarding topology. A rotation for an algorithm has been developed for the recognition of features and to identify important rotation features such as cylindrical, conical and toroidal surfaces. Holes in the rotating part are also created in the radial and axial directions by the identification algorithm. Algorithms recognize special convoluted features such as developing threads. Specific feature types are identified in the sub model's STEP file by finding their geometric/topological configuration. 3D CAD systems can easily recognize features using advanced algorithms.

MazinAl-wswasi1,2 · Atanas Ivanov1 [2] prove that the data of the topology and geometrical of product are arranged according to a hierarchical structure. The shell, which is a topological element, delimits the area in three-dimensional space connecting faces with edges. Subject line 253 is contained in the data section of STEP file AP203. The physical text file does not reflect the hierarchical arrangement of the data in the STEP file. Closed habitat nodes are not necessarily at the beginning of the hierarchy, but represent the highest level of organization. Data sections can be specified in any order, beginning or ending. Based on this statement, the data at all levels is considered reliable.

Muhammad Ali Kiani and Hassan Saeed [3] set the STEP file to read-only in your local workspace and save it to a locally saved file. These steps will be taken. To save the table, save the file separated by semicolons. A shell-based surface model needs to be validated. In this line number you will find the hashtag. If true, count the number of hashtags on this channel. The total number of welding points is determined by the project. Open Shell's advanced features are achieved by parsing files and concatenating their tokens. Advanced Surface calculates hash-tag numbers that provide pointers to the outer boundaries of surfaces and planes. Advanced Surface calculates outer boundary indicators for surfaces and planes using hashtag numbers. In the next step, outline the border using the hashtag from the outer edge of the face. Plan details will be accepted. Use the edge loop hashtag to get information about related edge loops. This process continues.

Xuesong Guan, Guangwei Meng, Xiuhua Yuan [4] According to them, feature classification is necessary for robust models using neural networks because neural networks are used for feature recognition. Various artificial neural network algorithms are used in the recognition process. The input layer in the basic ANN architecture consists of several hidden layers and neurons, and one or more neurons form the output layer. In this paper, we use a multilayer BP neural network with hidden layers to create a feature recognition function. In this work, a multilayer BP neural network with hidden layers is used to create a feature recognition function. The number of neurons in the output layer is equal to the number of recognized features. The hidden layer contains 10 neurons and the input layer contains 25 neurons. Slots, steps, holes and blind holes are called slots, steps, holes and blind holes.

Adem ÇİÇEK and Kubilay ASLANTAŞ [5] the elements are represented by a square matrix called the characteristic matrix. An object's geometry is defined by several surfaces. This function determines the matrix size in terms of rows and columns. The size of rows and columns in the object matrix is determined by the surfaces that make up the object's geometry. A machine object is represented by a series of objects that model the topological and geometric aspects of the part. The surfaces that make up the part's elements are defined by a topology that describes the affinity relationships. The system divides these relations into three groups: concave, convex and discontinuous. These surfaces are only adjacent if there is a connection between them. An array of elements representing numeric values that indicate contigs. A cell in an object matrix appears when the neighbour relationship between two adjacent faces in the object is convex and convex.

Victoria Miles1 ·Stefano Giani1 ·Oliver Vogt [6] create a STEP parser that interprets raw data from a STEP file and turns it into a structured data tree. This tree serves as input to the recursive encoder. Each line of file corresponding is represented by a node, and additional nodes are created to store information about coordinate points. The tree structure is reduced and unnecessary nodes are filtered out. A network of recursive encoders was developed to analyse the processed data tree. The Child Sum Tree Encoder-LSTM is the backbone of the network. It works through a hardcoded tree structure and uses in Recursive Coding Networks, a set of weights is learned on each node. This is how the weights are described in each. sections. Machine performance classification problems are used to evaluate the performance of recurrent encoder networks and demonstrate the feasibility of the proposed approach.

Thomas, Christophe, Samir Lamouri [7] states that previous files of AP203 were analysed and heuristics were built to identify them. Find the definition of a product formed in the form of a spiral sweep. Taking objects with closed shells and connected faces from a high-level B-rep shape representation creates multiple B-rep solids. The surface properties of the helical reamer are determined by the shape of the surface and the curvature of its edges. If the geometry is flat, the four sides will be curved. The spiral path object type is determined as a square based on the order of its edges. If the surface of the screw remover is spherical, the thread feature is identified as impact. Heuristics are designed to identify other types of helix substitution features and identify related helix substitution features.

Bitla and Venkateswara Rao [8] have developed fire prevention system works mainly with flat surfaces. Non-planar surfaces, including base surfaces and boundaries, are identified by the system. The non-planar surfaces mentioned above are present in the design of fireproof systems. The developed FR system performs feature recognition in two steps. The AP203 format represents and extracts geometric and topological information from the surfaces and edges of originally viewed objects. In the second step, the algorithm identifies and represents elements along non-planar surfaces for use in process planning according to the AP224 format.

Bitla, Venkateswara and Deepanshu [9] proposed that various Java classes, including user-defined classes, have been programmed to develop a feature recognition system that identifies B-spline surface features. The private wrapper object for the private wrapper variant is derived from the high-level B-rep form representation of the JSDAI model. A solid model consists of a set of connected faces. All instances of closed shell surface elements are retrieved and checked to see if they are boundary surface instances.

3. Challenges:

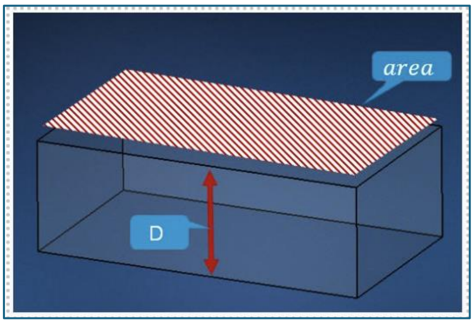

Fig 3: 3D model with axis and area

3D models are becoming more and more complex and CAD technology is developing rapidly. 3D geometry data is increasingly being analysed and interpreted in design and manufacturing, increasing the need for automation tools. [12]. The format of ISO 10303 express language STEP is a widely accepted standard for representing 3D product models. The problem of identifying and classifying objects in files lies in the many geometric shapes and the complexity of the set.

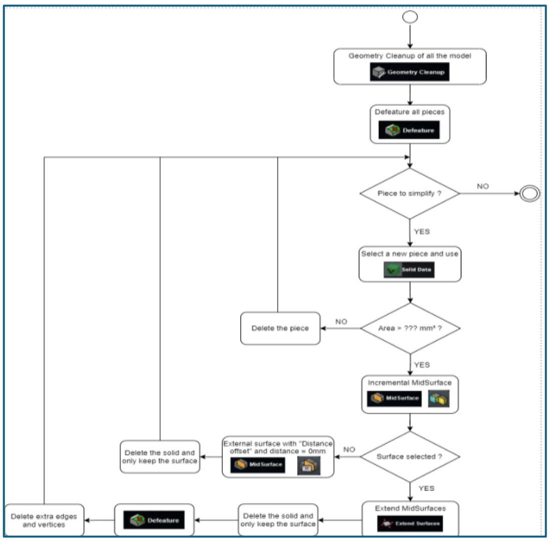

Fig 4: Challenges in identifying STEP elements

Despite its widespread use in industry, STEP files are not effectively and efficiently recognized or classified based on their geometric properties. Complex formats, different functional representations and large sizes of STEP files often pose challenges to existing methods. [10]The research project aims to develop new algorithms and methods for accurate, efficient and scalable geometric feature recognition from 3D models in STEP files.

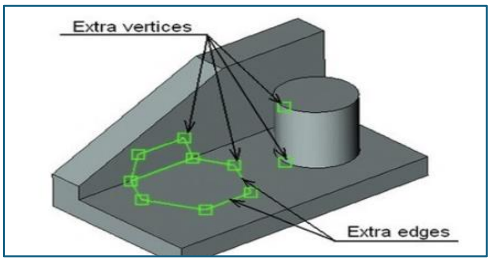

Fig 5: STEP model with edges and vertices

From simple to complex, new algorithms can be developed and explored to accurately identify and classify various geometric features in STEP files. (e.g. holes, slots, fillets, chamfers, free-form surfaces). [13] You can learn to handle and recognize complex geometries and features in large assemblies to solve problems such as feature interactions. You can also handle complex geometries and large assemblies and identify features through procedural exploration. Issues with resource interactions have been fixed. The geometric data in the STEP file is examined to determine the best way to analyse, simplify, and analyse this data to determine the correct elements. [14]. It goes like this: The best way to parse, simplify, and analyse the geometric data in a STEP file is to define the objects correctly. This process is done by checking the data in the STEP file. The solution must be designed to handle large STEP files containing multiple components and assemblies and provide an efficient identification process. Machine learning and artificial intelligence techniques can be integrated to improve the ability to recognize features, especially complex or ambiguous geometric shapes. [14]. This can be achieved by leveraging the system's potential to use features recognize with the help of machine learning and artificial intelligence. A framework needs to be established to evaluate proposed methodology and performance, accuracy against existing approaches using different sets of STEP files. This means that a framework is created and used to evaluate the functionality of a method by comparing it to existing methods. The evaluation is based on different STEP files, each with specific characteristics.

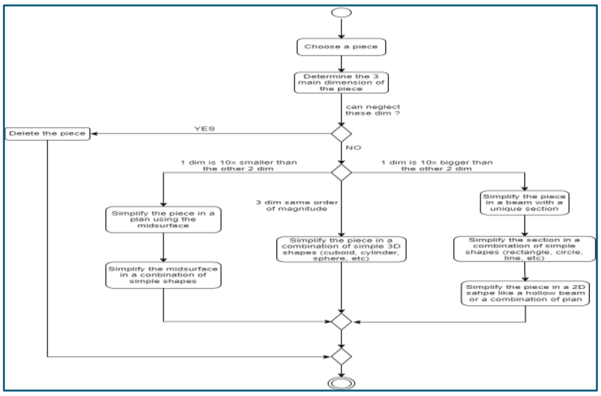

Fig 6: Flow of fetching object properties in 3D STEP

4. Proposed Methodology:

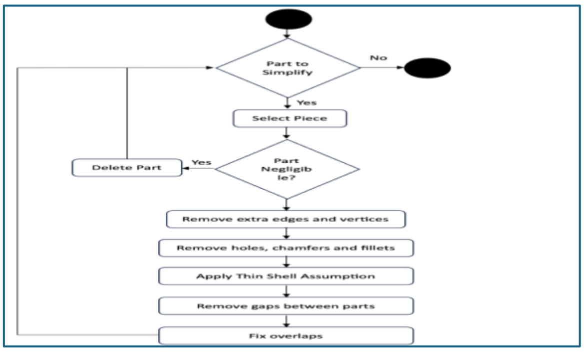

Fig7: Flow of proposed model

4.1 Parse STEP File:

Geometric and topological data from STEP 3D files are extracted and interpreted during the analysis process. The STEP file format, based on the ISO 10303 standard, is understood and used by specialized libraries or tools in this process. [16]The contents of the STEP file are parsed and its headers are read to retrieve the file's metadata. Metadata is identified and extracted. Entity definitions are also identified and retrieved. Data structures are created that represent geometric elements and interpret topological relationships between objects. A 3D model is then created based on the analysed data.

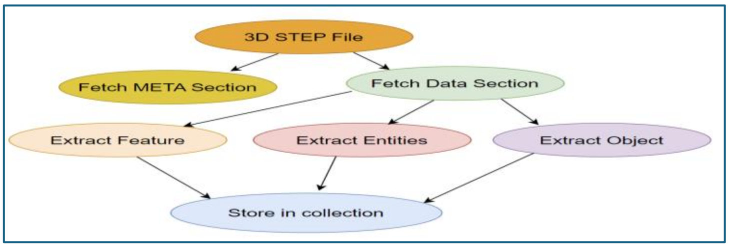

Fig 8: Sequence involve in parsing STEP file

Header section: Contains metadata about the file, such as file name, organization, author, and the default STEP version used. Data section: list of model objects containing geometric and topological data. Each object has a specific type, such as VERTEX_POINT, EDGE_CURVE, FACE_BOUND, etc. Several libraries and tools can analyse STEP files. Common options include:

1.PCC
2. Mesh
3. CadQuery
4. Pyspark.

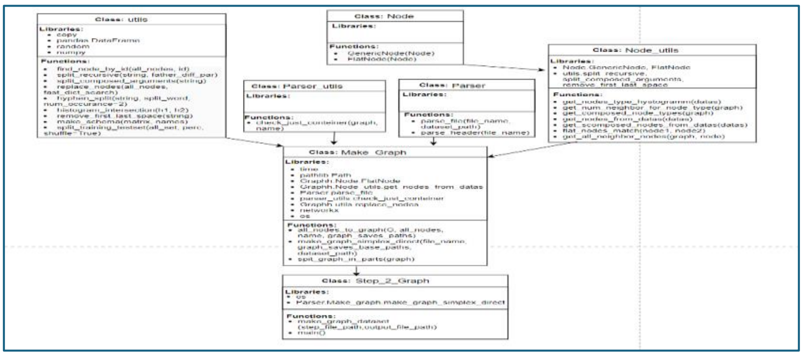

Fig 9: UML model of STEP parsing

4.2 Development of Datasets:

STEP files are collected or generated to create a dataset. You can use metadata to organize them and tag them for specific tasks, such as identifying sources. [13]You need to create an organized, documented and ready-to-use 3D STEP file dataset for your research, development or machine learning tasks. You can extend this dataset by adding more complex models, annotations, and metadata.

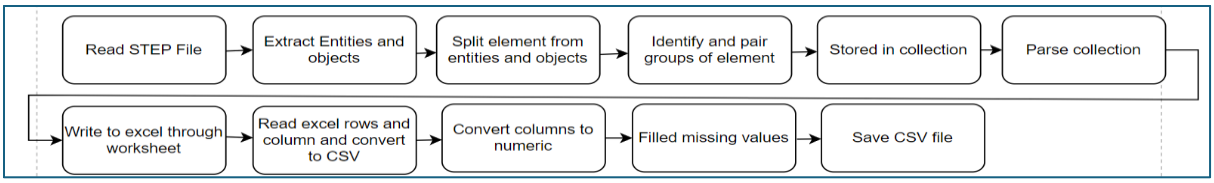

Fig:10 Sequence step for creating dataset

3D models are compiled in STEP format with different types covering different levels of complexity and different types of parts and assemblies.

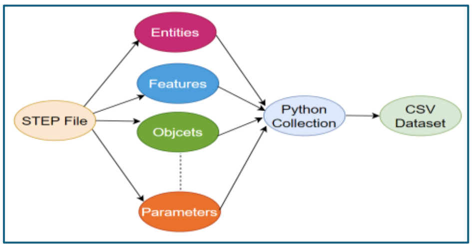

Fig:11 STEP file objects fetching for dataset

File headers are analysed to get basic information. This information includes file size, creation date and creation system. Get entities and objects from the data part of the file. Each element is stored in the collection with its key value during the descending process. Elements are identified manually or semi-automatically as geometric shapes. They are indicated by holes, cracks, depressions, protrusions, etc. You can create projections or 2D sections. You can calculate basic statistics such as volume, area and bounding box. Store metadata and annotations in a structured format such as JSON and XML

KEY	ENTITY	D0	D1	D2	D3	D4	D5	D6	D7	REFERENCES
1	CARTESIAN_POINT	0	26.5	-27.8	55	0	0	0	0	[]
2	PRODUCT	0	0	0	1148	0	0	0	0	[1148]
3	PERSON	0	0	0	0	0	0	0	0	[]
4	CIRCLE	0	1399	17.5	0	0	0	0	0	[1399]
5	EDGE_LOOP	0	1820	1815	0	0	0	0	0	[1820, 1815]

Table 1: Sample of 3D STEP dataset

4.3 Dataset Cleaning:

Dataset cleaning identifies and corrects errors and inconsistencies, enabling better analysis. Errors and inconsistencies are identified and corrected during the dataset cleaning process to improve the analysis. Once cleaned, the dataset is prepared for modelling, analysis, or other tasks to ensure accuracy and consistency. [17]3D Geometry STEP files undergo data cleansing to identify and correct issues that prevent proper interpretation, analysis, or processing of the model. STEP files are guaranteed to be free of errors, inconsistencies and unnecessary data during processing. STEP 3D files are checked to eliminate geometric and topological errors, simplify the model when necessary, and ensure metadata consistency during the cleanup process. Tools play an important role in automating processes and providing data sets for analysis and other downstream tasks.

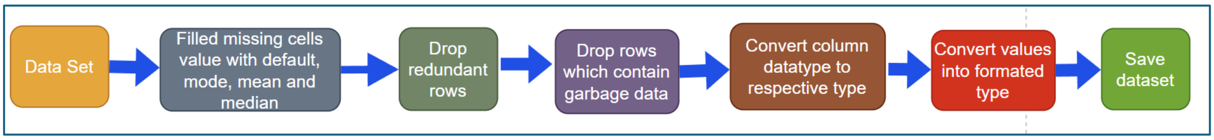

Fig 12: Dataset cleaning process

This STEP file needs to be checked for readability and corruption. Compliance with the ISO 10303 standard is essential. Remove unnecessary geometric elements from your design. Combine repeated or similar features to simplify the composition.

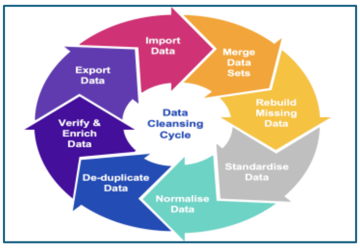

Fig 13: Cleaning cycle for datasets

Surfaces must be modified to eliminate holes, overlaps or self-cuts. Conflicts in personal relationships must be resolved. Convert all models to a single unit system to adjust geometric tolerances across files. Or from: The geometric tolerances of all files are synchronized, and all models are converted to a single unit system. Highly detailed elements can be simplified while maintaining the overall shape. You can remove unnecessary small elements and artifacts. Naming conventions for objects and attributes must be standardized. Incomplete or incorrect header information must be completed or corrected.

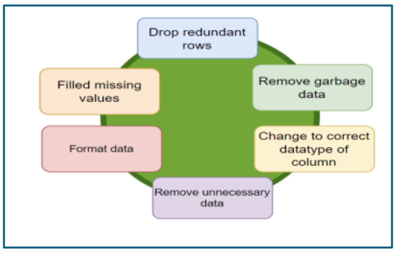

Fig 14: Process involves for formatting dataset

4.4 Perform EDA:

A set of STEP 3D geometry files were analysed using exploratory data analysis (EDA) to identify patterns, trends and insights that may be useful for further exploration and model creation. In connection with 3D geometric data, traditional research data analysis shifts from purely numerical and categorical data analysis to statistical and geometric/topological analysis. [18] STEP 3D geometric data sets were subjected to statistical, geometric and topological data analysis. The structure and complexity of a data set can be understood by examining and interpreting its central geometric features, visualizing patterns and identifying relationships. This process lays the foundation for understanding the data and paves the way for deeper analyses, feature development or model development.

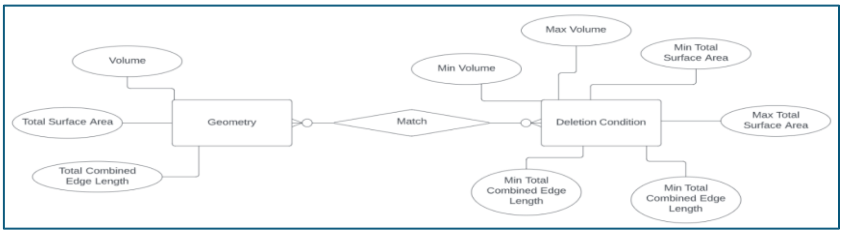

Fig 15: Process for exploratory data analysis EDA for 3D STEP

Count the total number of files. The average file size is calculated. Defines the creation date distribution.

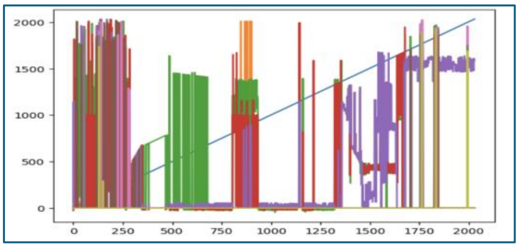

Fig 16: Bar chart for 3D STEP data

Volume distribution and surface area were analysed. The bounding box dimensions are marked. Proportion and symmetry measurements are calculated.

DESCRIPTION	KEY	D0	D1	D2	D3	D4	D5	D6	D7
Skew	0.000560	9.558666	1.067832	1.994086	1.137753	5.648787	12.027604	18.278033	22.38694
Correlation	0.000560	9.558666	1.067832	1.994086	1.137753	5.648787	12.027604	18.278033	22.38694
count	2030.0000	2030.0000	2030.0000	2030.0000	2030.0000	2030.0000	2030.0000	2030.0000	2030.0000
mean	1017.389163	17.757143	435.452905	262.397762	417.854768	53.789655	12.882266	5.615271	3.717734
std	587.860103	151.100114	584.338418	489.346093	651.047086	305.363280	150.594077	100.836957	80.830690
min	1.000000	0.000000	-47.000000	-27.750000	-1.000000	0.000000	0.000000	0.000000	0.000000
25%	508.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	1016.500000	0.000000	23.750000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000
75%	1526.750000	0.000000	796.750000	399.750000	860.750000	0.000000	0.000000	0.000000	0.000000
max	2034.000000	2012.000000	2032.000000	2034.000000	2033.000000	2022.000000	2025.000000	1956.000000	1932.000000

Table 2: Dataset description table

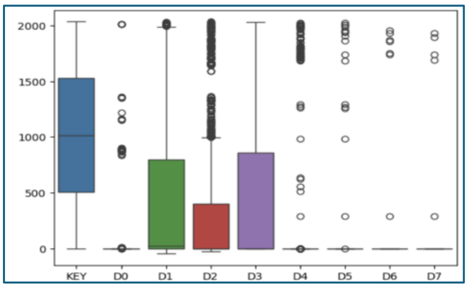

Fig 17: Swarn plot for 3D STEP data

Geometric features are classified and calculated in the model. The distribution of these signals was analysed. Correlations between signals were revealed.

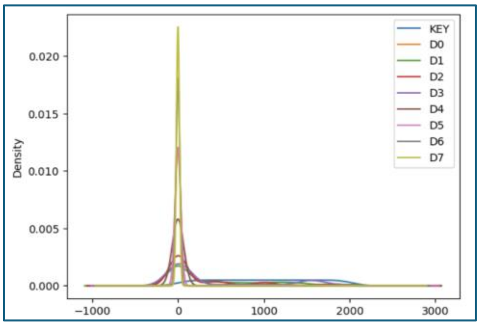

Fig 18: Skew plot for 3D STEP data

The number of faces, edges and vertices is checked. Connection and connection are analysed. Types of surfaces (flat, cylindrical, arbitrary shape) were investigated. Controls the number of faces, edges and vertices in the model. Connection and connection are analysed. Types of investigated object surfaces (plane, cylinder, surface of arbitrary shape).

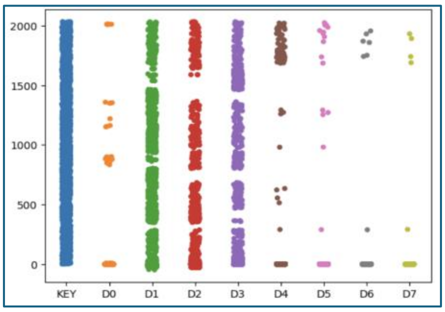

Fig 19: Strip plot for 3D STEP data

Use the difficulty level and analyse its distribution. Determine the simplicity or complexity of your model.

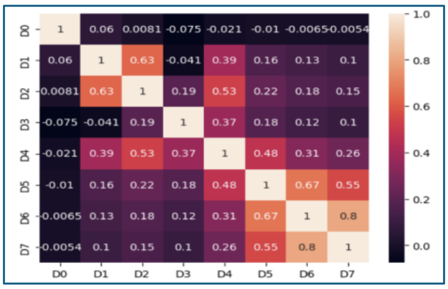

Fig 20: Heap matrix 3D STEP data

You can create a histogram of key indicators. Scatterplots can be prepared to show relationships. A three-dimensional representation of a typical model is possible.

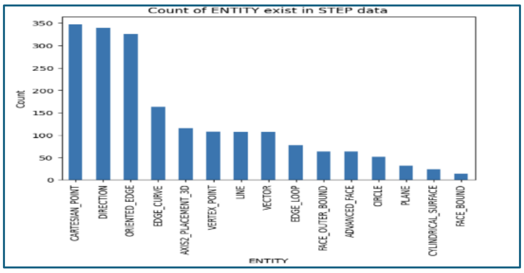

Fig 21: Entity frequency data

4.5 Encode Trained data with ML:

Machine learning algorithms process encoded 3D geometric data into a usable format. Object recognition, feature detection, classification and creative modelling are essential to this process. The ML model, the nature of the problem, and the type of geometry will determine which coding method to use. [10] Machine learning models process 3D geometric data after encoding it into the appropriate format. The specific application, data characteristics, and available computational resources determine the encoding choice. By choosing and using these encoding methods carefully, you can maximize the potential of 3D data for a variety of machine learning tasks.

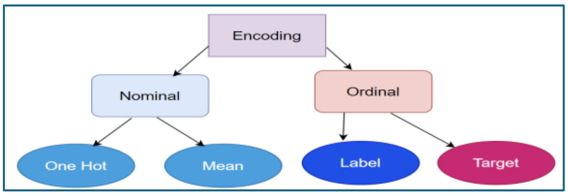

Fig 22: Data encoding process

You can generate voxels by converting a 3D model to a mesh. This process uses binary or multivalued voxels to represent the occupancy of each grid cell. The points are represented in the model with three-dimensional coordinates and additional features. The sample 3D coordinates and additional features from model surfaces. Transforms the model representation to create a graphical structure. Replace faces, edges and vertices with nodes. Determine the connections between these nodes.

Volume	Center of Mass	Inertia Matrix	Surface Area	Number of Faces	Number of Solids	Number of planes	Number of cylinder
76615.50864724086	(2.914442331989318, -3.4739588334497657, 26.815249660078138)	[[42810477.086416915, 3572053.588912038, -49351.35669452418], [3572053.588912038, 37408446.957851, -801419.2423244929], [-49351.35669452418, -801419.2423244929, 41357857.946387134]]	22800.12557970294	1	1	0	1

Table 3: 3D STEP file geometrical properties

For each model, geometric and topological properties are calculated. For each model, a feature vector of fixed length is created. [11]. You can create multiple 2D views from a 3D model. Projection is available for each image-based encoding. Multiple 2D views can be created for a 3D model. An autoencoder is used to learn the compressed representation and encode the model into a low-dimensional latent space. The model is encoded in a low-dimensional latent space using an autoencoder that learns a compressed representation.

4.6 Process of Machine Learning ML:

Machine learning (ML) processes 3D STEP geometric data using a multi-step sequence. This sequence includes data preprocessing and coding followed by model training and evaluation.

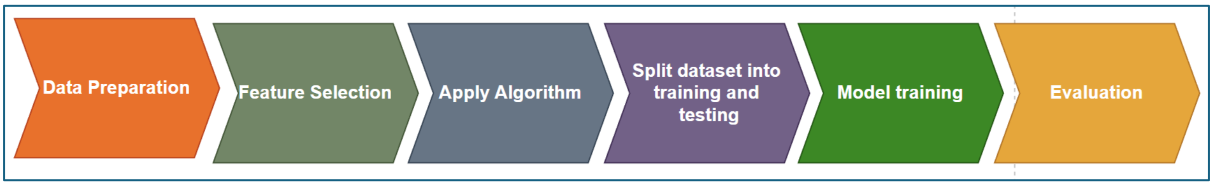

Fig 23: Process step sequence of machine learning

Encoding can be applied to STEP files to generate voxels or point clouds. The data from this coding process must be normalized or standardized. The corresponding geometric and topological features are revealed. [7]Dimension reduction methods are used if necessary. You can choose the appropriate algorithm depending on the task. For example, CNN's are suitable for voxel-based tasks such as classification. Practice, test and test. Custom normalization or standardization must be used. Data is formatted for training by putting it in an appropriate format such as a tensor. The data set is then divided into individual sets.

Face	Area
1	110.58 square units
2	363.89 square units
3	110.58 square units
4	363.89 square units
5	78.54 square units
6	26.39 square units
7	8.08 square units
8	78.54 square units
9	26.39 square units
10	8.08 square units
11	75.34 square units
12	75.34 square units
13	77.71 square units
14	880.00 square units
15	84.82 square units

Table 4: Geometrical properties of face and area in STEP file

To evaluate the performance of your model, you need a validation set and a test set. Metrics such as accuracy, precision, recall, and F1 score must be calculated based on the task given when test data is presented to the model. The performance of the model must be verified using validation and test sets. Calculations should be made for accuracy, precision, recall, F1 score, and other metrics appropriate to the task at hand. The test data is analysed and the model calculates the metric after receiving it.

4.7 Data fitting using CNN:

Integrating 3D STEP data into CNN (convolutional neural network) is challenging due to the 3D nature of the data and the unique 2D input structure of the CNN. 3D STEP data is difficult to process with a convolutional neural network (CNN) due to the 3D nature of the data and the default 2D input structure of the CNN. After loading and preprocessing the data, a CNN architecture is created to fit the training data. Access model performance to make predictions. [19]The data is loaded and pre-processed. A CNN architecture is built and given training data. Access model performance to make predictions. The ability to convert 3D data into an appropriate format allows multiple steps to be performed, such as data preprocessing, model training, and evaluation.

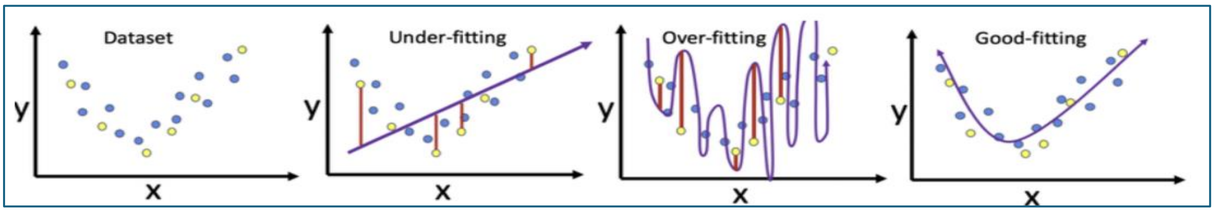

Fig 24: Fitting graph with CNN model

3D data must be in an appropriate format: a series of 2D images (frames) or a 3D matrix. You can normalize your data to optimize the learning process. You can use data augmentation techniques to augment different types of training data. A small cube in each frame indicates whether any part of the object occupies that space.

A 3D object can be thought of as a set of 3D points that describe its shape when viewed as a point cloud. [20]The surface of a 3D object generates a series of points. By normalizing a point cloud, you can create a point cloud that corresponds to a unit sphere or cube. This can be achieved in various ways, for example by scaling the coordinates or transforming the data using certain techniques. Once normalized, the point cloud is processed for further analysis using specialized network architectures such as point networks.

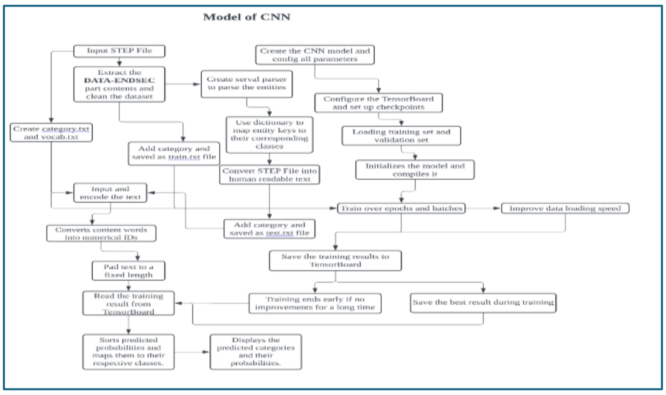

Fig 25: Flow model of fitting data into CNN

Regardless of the format chosen, CNN requires data in the correct format and preprocessing is required. Adjust and resize 3D coordinates within a specified area. The training data undergoes transformations, including rotation, translation, and scaling, to increase in diversity. The transformations, such as rotation, translation, and scaling, are applied to the training data to increase its diversity.

This approach uses a 3D convolution layer (Conv3D) to incorporate all three dimensions of the data. You can add a 3D pooling layer (MaxPooling3D) to reduce the spatial size of your neural network. A ReLU activation function can be introduced to introduce non-linearity. A 3D pooling layer (MaxPooling3D) is added to the neural network to reduce the spatial dimension. The ReLU activation function is introduced to introduce non-linearity. Fully connected layers can be added at the end of the network for classification or regression tasks.

Avoid overfitting by having enough training data. 3D data in a data format must be processed with an appropriate batch size. Make sure you have enough space in your training to avoid overtraining.

4.8 Apply CNN:

A CNN (convolutional neural network) processes 3D geometric STEP data. or 3D STEP geometry data is processed using a CNN (convolutional neural network). You can create and train a CNN model to transform 3D data into an appropriate format. Convert and preprocess 3D data into a format suitable for CNN models. [21]. This may include data normalization, dimensionality enhancement, etc. CNN must be able to handle data containing complex geometric and topological information. CNN typically need to convert STEP data into usable formats, such as voxel grids, point clouds, or Multiview projections, before using them. The CNN architecture can vary depending on the data format used (voxel, point cloud or Multiview image).

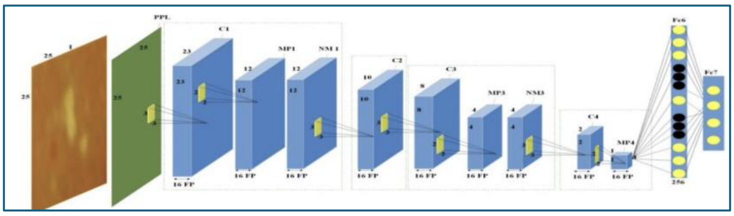

Fig 26: CNN process flow model

Machine learning models are considered, both traditional machine learning algorithms and large language models. Different approaches and criteria for processing 3D geometric data depend on the model type and the specific task. The type and specific tasks of the model determine the approach and quality of 3D geometric computing. The specific task of evaluating ML and LLM models based on 3D geometric data requires understanding and adapting evaluation metrics. Traditional machine learning models prioritize geometric accuracy and data accuracy in their approach. However, when using a larger language model, the emphasis is on the quality and consistency of the resulting text or code.

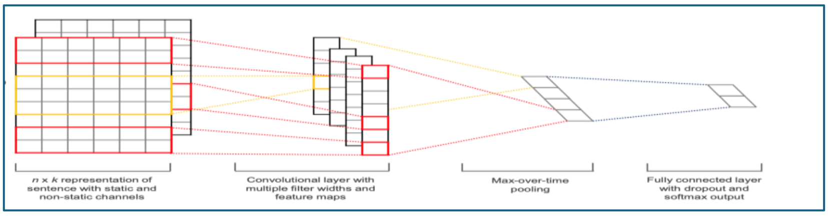

Fig 27: CNN Layer transformation process

We need to measure the percentage of correct predictions. The balance between precision and recall, as well as performance evaluation between levels, can be described in the precision and recall are evaluated together in their balance. Performance analysis is performed on several classes. Regression problems require measuring and analysing the performance of different segments. We need to calculate the root mean square of the difference between the predicted and actual values.

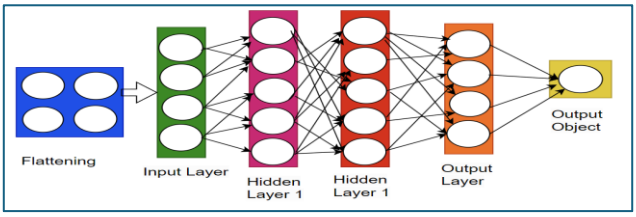

Fig 28: CNN filtering layer process with hidden layer

Use K-fold cross-validation to ensure that the model generalizes well. Model parameters can be found using methods such as web search and random search. Compare the performance of this model with the reference model. Learn about the performance of advanced models such as 3D CNN and PointNet.

They focus on quantitative measures such as is a measure of the number of true positives out of all the instances like precision, accuracy, recall, and root mean square error. This metric is used depending on the task. Standard techniques such as cross-validation, hyperparameter tuning and visualization are used. [22]. Depending on the task, quantitative indicators such as accuracy, reliability, completeness and root mean square error are used. Standard techniques such as cross-validation, hyperparameter tuning and visualization are used. These indicators, such as BLUE, RED and F1 scores, are language-related and used for evaluation. Estimates for specific tasks are often based on human judgment.

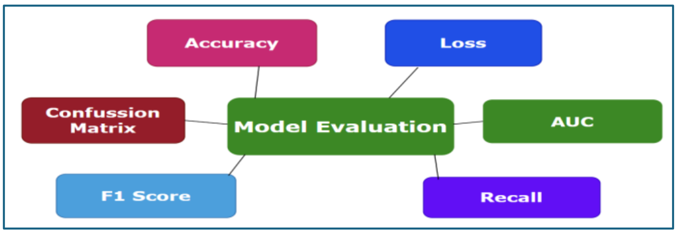

Fig 29: Model evaluation process

5. Result and Calculation:

5.1 Random Forest:

Model accuracy score with 10 decision-trees: 0.4507
Model accuracy score with 100 decision-trees: 0.4507

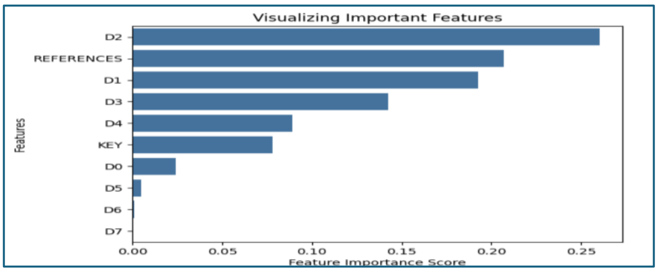

Fig 30: Visualize of entities and feature

entities	precision	recall	f1-score	support
accuracy			0.45	670
macro average	0.17	0.21	0.18	670
weighted average	0.35	0.45	0.37	670

Table 5: Random forest accuracy score

Null accuracy score: 0.7439

classification accuracy	classification error	precision	Recall or Sensitivity	specificity
(TP + TN) / (TP + TN + FP + FN)	(FP + FN) / (TP + TN + FP + FN)	TP / (TP + FP)	TP / (TP + FN)	TN / (TN + FP)
0.3453	0.223	0.7463	0.3231	0.5397

Table 6: Random forest classification result

ROC AUC	Cross validated ROC AUC	Cross-validation scores	Average cross-validation score
0.5659	0.35255	[0.44535277 0.6565231 0.34237692]	0.5544

Table 7: Random forest ROC and cross validation

5.2 Logistic Regression:

Model accuracy score: 0.4655
C=1
Training set score: 0.6743
Test set score: 0.5443

entities	precision	recall	f1-scor	support
accuracy			0.54	406
Macro average	0.18	0.18	0.17	406
Weighted average	0.43	0.54	0.45	406

Table 8: Logistic regression accuracy score

Null accuracy score: 0.7759

classification accuracy	classification error	precision	Recall or Sensitivity	specificity
(TP + TN) / (TP + TN + FP + FN)	(FP + FN) / (TP + TN + FP + FN)	TP / (TP + FP)	TP / (TP + FN)	TN / (TN + FP)
0.3453	0.654	0.8763	0.3451	0.8797

Table 9: Logistic regression classification result

ROC AUC	Cross validated ROC AUC	Cross-validation scores	Average cross-validation score
0.2359	0.76545	[0.44346277 0.6462231 0.3235492]	0.6575

Table 10: Logistic regression ROC and cross validation

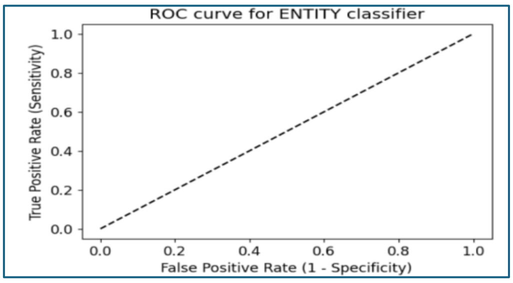

Fig 31: Visualize ROC curve and entity classifier

5.3 Decision Tree:

entities	precision	recall	f1-score	support
accuracy			0.96	609
macro average	0.56	0.6	0.57	609
weighted average	0.96	0.96	0.96	609

Table 11: Decision tree accuracy score

Null accuracy score: 0.7549

classification accuracy	classification error	precision	Recall or Sensitivity	specificity
(TP + TN) / (TP + TN + FP + FN)	(FP + FN) / (TP + TN + FP + FN)	TP / (TP + FP)	TP / (TP + FN)	TN / (TN + FP)
0.6143	0.728	0.9832	0.5631	0.7897

Table 12: Decision Tree classification result

ROC AUC	Cross validated ROC AUC	Cross-validation scores	Average cross-validation score
0.3653	0.4336	[0.4325677 0.6453521 0.3456752]	0.3562

Table 13: Decision Tree ROC and cross validation

5.4 Naive Bayes:

Model accuracy score: 0.0870
Training-set accuracy score: 0.9289
Training set score: 0.9289
Test set score: 0.0870

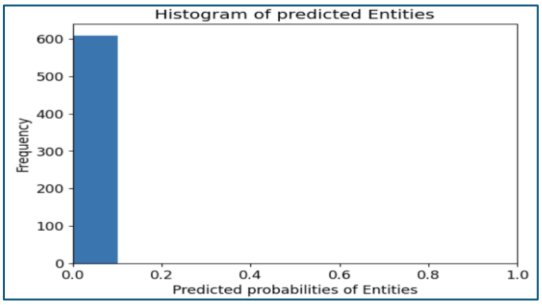

Fig 32: Visualize predicted entities

entities	precision	recall	f1-score	support
accuracy			0.09	609
macro average	0.03	0.09	0.04	609
weighted average	0.02	0.09	0.03	609

Table 14: Naive Bayes accuracy score

Null accuracy score: 0.8549

classification accuracy	classification error	precision	Recall or Sensitivity	specificity
(TP + TN) / (TP + TN + FP + FN)	(FP + FN) / (TP + TN + FP + FN)	TP / (TP + FP)	TP / (TP + FN)	TN / (TN + FP)
0.3464	0.345	0.7564	0.2345	0.3543

Table 15: Naive Bayes classification result

ROC AUC	Cross validated ROC AUC	Cross-validation scores	Average cross-validation score
0.5543	0.3245	[0.4235377 0.6435621 0.3865752]	0.4352

Table 16: Naive Bayes ROC and cross validation

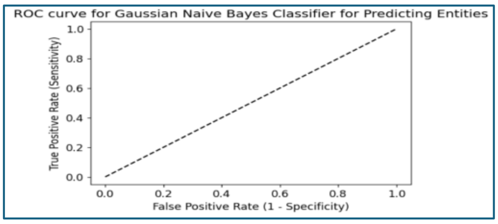

Fig 33: Visualize ROC curve for naive bayes

5.5 K-Nearest Neighbour KNN:

Model accuracy score: 0.9310
Training-set accuracy score: 0.9606
Training set score: 0.9606
Test set score: 0.9310

Model accuracy score with k=5: 0.9261

entities	precision	recall	f1-score	support
accuracy			0.91	406
macro	0.5	0.54	0.51	406
weighted	0.9	0.91	0.9	406

Table 17: KNN accuracy score

Null accuracy score: 0.7549

classification accuracy	classification error	precision	Recall or Sensitivity	specificity
(TP + TN) / (TP + TN + FP + FN)	(FP + FN) / (TP + TN + FP + FN)	TP / (TP + FP)	TP / (TP + FN)	TN / (TN + FP)
0.3456	0.3246	0.6554	0.5653	0.3434

Table 18: KNN classification result

ROC AUC	Cross validated ROC AUC	Cross-validation scores	Average cross-validation score
0.6553	0.2356	[0.2354677 0.4653421 0.326762]	0.9855

Table 19: KNN ROC and cross validation

5.6 Support Vector Machine SVM:

Model accuracy score with default hyperparameters: 0.7414
Training-set accuracy score: 0.8233
Training set score: 0.8233
Test set score: 0.8103

entities	precision	recall	f1-score	support
accuracy			0.81	406
macro average	0.58	0.58	0.57	406
weighted average	0.82	0.81	0.79	406

Table 20: SVM accuracy score

Null accuracy score: 0.344

classification accuracy	classification error	precision	Recall or Sensitivity	specificity
(TP + TN) / (TP + TN + FP + FN)	(FP + FN) / (TP + TN + FP + FN)	TP / (TP + FP)	TP / (TP + FN)	TN / (TN + FP)
0.2366	0.3464	0.3445	0.5543	0.6432

Table 21: SVM classification result

5.7 Convolutional Neural Network CNN:

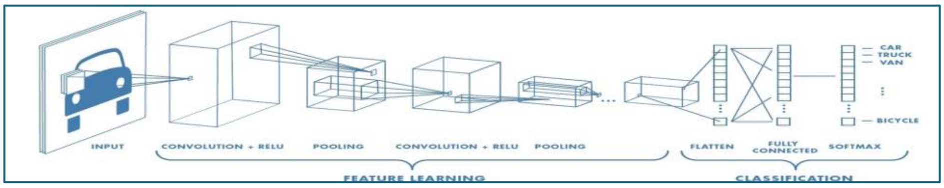

Fig 34: Visualize view of Convolutional Neural Network CNN

Layer	**Output shape**	**Parameter**
conv2d	(None, 30,34,12)	345
max_pooling2d	(None, 15,41,38)	34
conv2d_1	(None, 11,63,21)	564
max_pooling2d_1	(None, 6,23,56)	21
conv2d_2	(None, 33,15,18)	357
max_pooling2d_2	(None, 15,6,33)	765
flatten	(None,18,28,35)	45
dense	(None, 31,22,17)	356
dense_1	(None, 11,24,31)	31

Table 22: CNN layer parameters and shape

Total parameters	325,741
Trainable parameters	450,960
Testing parameters	35,740

Table 23: CNN training and testing parameters

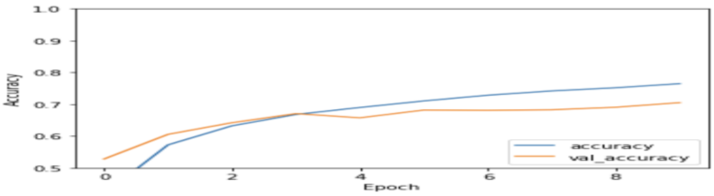

Fig 35: Graph of accuracy for CNN model

Test Accuracy is 0.4534654263214111

```
Time usage: 0:00:00
Training and evaluating...
Epoch 1/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 3s 7ms/step - accuracy: 0.0991 - loss: 3.4355 - val_accuracy: 0.1627 - val_loss: 2.6975
Epoch 2/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1658 - loss: 2.8845 - val_accuracy: 0.1605 - val_loss: 2.8234
Epoch 3/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1307 - loss: 2.9708 - val_accuracy: 0.1627 - val_loss: 2.6147
Epoch 4/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1823 - loss: 2.9035 - val_accuracy: 0.1627 - val_loss: 2.7285
Epoch 5/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1245 - loss: 2.9668 - val_accuracy: 0.1627 - val_loss: 2.6455
Epoch 6/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1670 - loss: 2.9353 - val_accuracy: 0.1627 - val_loss: 2.6403
Epoch 7/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1501 - loss: 3.0354 - val_accuracy: 0.1605 - val_loss: 2.6149
Epoch 8/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1218 - loss: 3.0140 - val_accuracy: 0.1605 - val_loss: 2.6597
Epoch 9/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1359 - loss: 2.8676 - val_accuracy: 0.1627 - val_loss: 2.8095
Epoch 10/10
125/125 ━━━━━━━━━━━━━━━━━━━━ 1s 4ms/step - accuracy: 0.1268 - loss: 2.9537 - val_accuracy: 0.1627 - val_loss: 2.6531
```

Fig 36: CNN result with 10 epochs

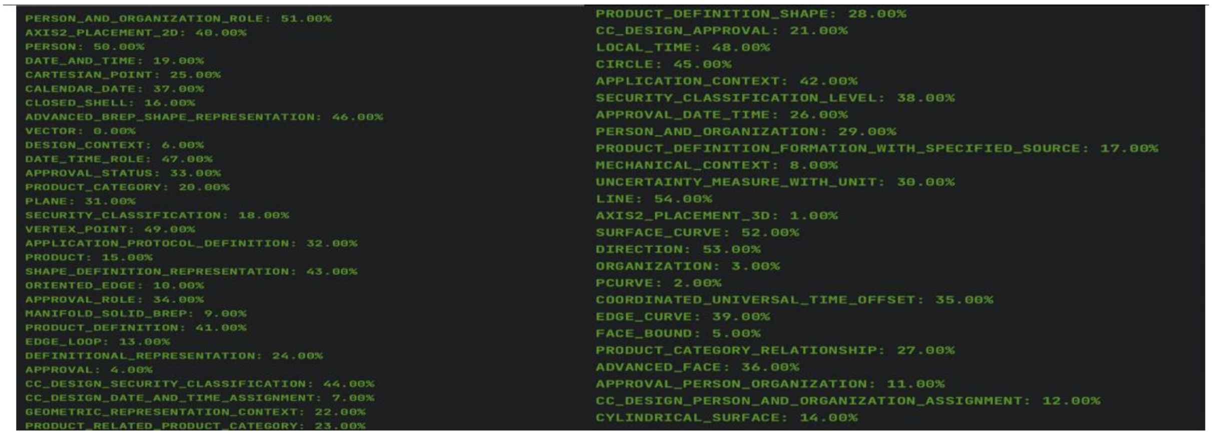

Fig 37: CNN result for entities score

7. Analysis and Discussion:

Algorithms	precission	recall	f1-score	support	weighted	accuracy	error
Random Forest	0.64	0.454	0.23	0.18	0.17	406	0.4354
Logistic Regression	0.47	0.731	0.45	0.55	0.12	432	0.5344
Decission Tree	0.23	0.635	0.64	0.61	0.21	124	0.8419
Naïve Bayes	0.87	0.763	0.17	0.34	0.28	675	0.4675
SVM	0.49	0.964	0.63	0.65	0.15	356	0.7664
KNN	0.25	0.562	0.74	0.32	0.19	754	0.8654
CNN (Convolutional Neural Network)	0.62	0.438	0.49	0.11	0.22	345	0.4554

Table 24: Algorithms accuracy, weighted and score comparison

Key performance indicators are evaluated to summarize the effectiveness of the algorithm. Depending on the type of task (classification), commonly used metrics are accuracy, precision, recall, F1 score, area under the curve (AUC), and mean square error (MSE). Trade-offs between algorithms are discussed. For example, an algorithm may have higher accuracy but require more computational resources. On the other hand, other algorithms may be faster, but they must evaluate each data set or part of the data. Algorithm robustness Even with a small amount of data, the accuracy of the model must be considered, as simple models such as neural networks may not provide satisfactory results.

Sr. No.	ML Algorithm	Cross validated ROC AUC	Cross Validation Score	ROC AUC Score
1	Logistic Regression	83.60%	89.16%	87.15%
2	Support Vector Classifier	82.20%	91.43%	83.64%
3	Decision Tree Classifier	81.84%	77.58%	79.24%
4	Random Forest Classifier	74.89%	95.65%	85.47%
5	K-Nearest Neighbours Classifier	88.25%	88.54%	71.35%
6	Support Vector Machine	84.32%	51.25%	75.32%
7	Convolutional Neural Network	77.14%	87.32%	49.36%

Table 25: All algorithms cross validations and ROC comparison

This dataset can be used to effectively understand binary classification problems that combine numerical and categorical features. By providing information, subject matter experts (in this case devices and keys) can help you take the next step. During feature development, feature selection testing is performed before data scaling. Pre-test data processing can reveal data manipulation, but the results remain the same regardless of the order of the steps. I can't find a problem due to lack of reading on the subject. Understanding this problem is an important part of eliminating outliers, even if the outlier detection test is positive. Data becomes more expressive through visualization. Visualization makes it easier to understand information and test results and output. Hyperparameter tuning is not performed during simulation. The algorithm's performance can be improved. The satisfactory performance of the algorithm is achieved.

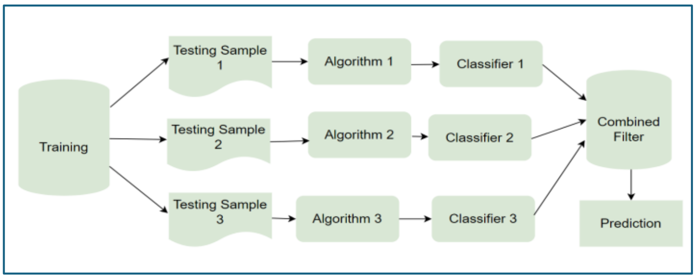

Fig 38: Flow of training and classifying multiple algorithms

Other models in our analysis consistently produced lower accuracy and F1 scores than the random forest algorithm, especially on noisy datasets. This comparison highlights the robustness of communication reuse methods described in the existing literature. Imbalanced datasets do not work well with SVM (support vector machines), but SVM work best when classes are clearly differentiated. This limitation is acknowledged by the researchers. With enough data, neural networks have great potential, but they are computationally expensive and prone to overfitting. With enough data, neural networks have high data processing power. However, it is computationally expensive and prone to overfitting. Smaller data sets may be recommended for further adjustment using normalization methods.

8. Conclusion:

In various applications such as CAD/CAM systems, manufacturing and quality control, STEP 3D geometry files are analysed to identify features. STEP files are processed to extract significant geometric and topological features for later use in classification, clustering, or modelling.

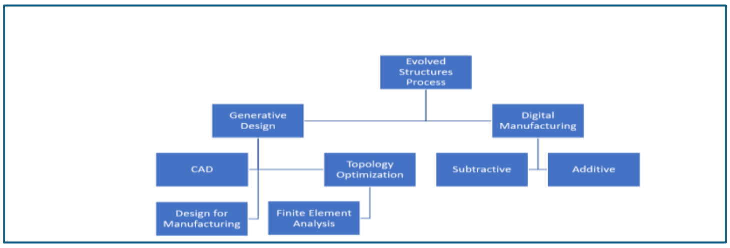

Fig 39: Flow of complete 3D STEP evaluation process.

Analysis of 3D geometry STEP files to identify features is essential in today's digital manufacturing workflow. Identifying the characteristics of a 3D geometric STEP file is a complex but important process, and an important step between raw 3D data and practical knowledge in various industrial applications.) In practice, this STEP file uses advanced machine learning techniques such as 3D CNN and PointNet to process data and ensure the accuracy of geometric features. Many CAD systems use machine learning techniques to improve accuracy and efficiency, especially when it comes to complex geometries and ensure compatibility between them. It is constantly improving and has great potential to improve design and manufacturing processes in the future.

References:

[1] D. Sreeramulu1 and C.S.P.Rao, "A new methodology for recognizing features in rotational parts using STEP," *International Journal of Engineering, Science and Technology,* vol. 3, no. 6, pp. 102-115, 2011.

[2] Mazin Al-wswasi and Atanas Ivanov , "A novel and smart interactive feature recognition system for rotational parts using a STEP file," *The International Journal of Advanced Manufacturing Technology,* vol. 104, no. 8, p. 261–284, 2019.

[3] V. Naga Malleswari, P. Lohith Raj and A. Ravindra , "Automatic feature recognition from STEP file for smart manufacturing," *Progress in Additive Manufacturing,* 27 03 2024.

[4] Muhammad Ali Kiani and Hasan Saeed, "Automatic Spot Welding Feature Recognition From STEP Data," in *2019 International Symposium on Recent Advances in Electrical Engineering (RAEE)*, 2019.

[5] Xuesong Guan, Guangwei Meng and Xiuhua Yuan , "Machining feature recognition of part from STEP file based on ANN," in *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE)*, 2010.

[6] Adem ÇİÇEK and Kubilay ASLANTAŞ, "3D Rendering USING FEATURE MATRICES RECOGNITION OF ITS ELEMENTS," *J. Fac. Eng. Arch. Gazi Univ.,* vol. 25, no. 1, pp. 147-155, 2010.

[7] Victoria Miles1, Stefano Giani1 and Oliver Vogt1, "Recursive encoder network for the automatic analysis of STEP files," *Journal of Intelligent Manufacturing (2023),* vol. 34, no. 26, p. 181–196, 2023.

[8] Baisheng Dai and Jinyou Qiao, "Research on STEP-based 3D model reconstruction and feature recognition," in *International Conference on Mechanic Automation and Control Engineering*, 2010.

[9] Bitla Venu and Venkateswara Rao Komma, "Some Heuristics of Helical Sweep Surfaces in ISO 10303 STEP AP203 File to Identify Threaded Features," in *International Conference on Advances in Mechanical, Industrial, Automation and Management Systems (AMIAMS)*, 2017.

[10] Bitla Venu and Venkateswara Rao Komma , "STEP-based feature recognition from solid models having non-planar surfaces," *International Journal of Computer Integrated Manufacturing,* vol. 30, no. 10, pp. 1-18, 2017.

[11] Bitla Venu, Venkateswara Rao Komma and Deepanshu Srivastava, "STEP-based Feature Recognition System for B-spline Surface Features," *International Journal of Automation and Computing,* vol. 15, no. 21, p. 500–512, 2018.

[12] Steven J. Owen,, Timothy Shead and Shawn Martin, "Defeaturing of CAD Models Using Machine Learning," *January 2024Journal of Advanced Manufacturing Systems,* vol. 23, no. 03, 2024.

[13] Birinderjit Singh Kalyan, "A multi-fidelity platform using machine learning for integrated design of aircraft systems," in *International Conference on Emerging Systems and Intelligent Computing (ESIC)*, 2024.

[14] Alaa M. Ubaid and Fikri T. Dweiri, "Sustainable systems integration by means of STEP-NC: Literature review," in *Advances in Science and Engineering Technology International Conferences (ASET)*, 2018.

[15] Hongyan Yang and Yanru Zhong, "Extracting Algorithm of Characteristics Data Based STEP in Rapid Prototyping Technology," in *Third International Conference on Genetic and Evolutionary Computing*, 2009.

[16] Deepanshu Srivastava and Venkateswara Rao Komma, "Development of STEP-NC for intelligent machining," in *International Conference on Advances in Mechanical, Industrial, Automation and Management Systems (AMIAMS)*, 2017.

[17] Thomas Paviot, Christophe Mouton and Samir LamouriAuthors, "Long term control of 3D engineering data for nuclear power plants," in *Web3D '13: Proceedings of the 18th International Conference on 3D Web Technology*, 2013.

[18] Donald T. Ward, David A. Redman and Bruce A. Lewis, "An approach to integration of complex systems: the SAVI virtual integration process," in *Proceedings of the 2013 ACM SIGAda annual conference on High integrity language technology*, 2013.

[19] Bingjie Yang, Enqing Chen, Shouyi Yang and Wenjuan Bai, "RGB-D geometric features extraction and edge-based scene-SIRFS," in *IEEE International Conference on Communication Software and Networks (ICCSN)*, 2015.

[20] Tobias Köhler, Buchao Song, Jean Pierre Bergmann and Diana Peters, "Geometric feature extraction in manufacturing based on a knowledge graph," *Heliyon,* vol. 9, no. 9, 2023.