Journal of Research Administration
Society of Research Administrators International

# ENABLING EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) TO NARRATE PATENT SCENARIOS IN DEVISING SCIENCE POLICY DECISIONS & TECHNOLOGY FORECASTS

**Zeba Hasan[1], M.Afshar Alam, Harleen Kaur, Ihtiram Raza Khan, Bhavya Alankar**

Department of Computer Science & Technology

School of Engineering Sciences & Technology-SEST

Jamia Hamdard, New Delhi, India

## Abstract

Machine learning methods, offering unique characteristics for variable identification and its importance, are now gradually occurring. The are now used for prediction, forecasting, and devising important scenarios for integrated decision-making. It shows that sequential and dynamic growth in emerging technologies can fundamentally become a foremost indicator to measure technology strength in innovation and development.

Similarly, prudent and systematic data generation is part of the evidence analysis and service/product development. Assisted machine learning solutions provide data-driven prediction for medical analysis, fraud detection & disease detection, etc. providing solutions with explainable intelligence. Assisting solutions with more data-driven explainable solutions, during the performance of advanced machine learning has created impactful stage for generating white box solutions out of the black box interpretation. Explainable artificial intelligence provides superior and solution-based AI, optimal solutions and an impactful decision-making solution. The approach of explainable artificial intelligence being used here represents the assessment for ML algorithms to predict accurate results. It is well represented that innovation and development can be identified via various innovation indicators, where patents are one of the important solution providers for futuristic technologies. In this paper patent data obtained from the Organisation for Economic Cooperation and Development (OECD) database, are used of patents for India have been taken to represent technological mandate.

**Key Words:** Analytics, Organisation for Economic Co-operation and Development (OECD), H2o.ai, Explainable Artificial Intelligence (XAI), Machine Learning, Decision Science

## Introduction

Increased availability of data has pushed ahead digital information access as well as technological ascendency to reduce the barriers and technology diffusion. Data-driven decision-making at the same, is becoming more prominent from a contemporary scenario of statistical modelling. Therefore, the aim of this study here is, to propose how to systematically put ahead *technology prediction*, by calculating the patent growth value. High technology development expenditure on information and communication technology R&D, has gradually been unveiled from 2015 to 2019.

In our study our objective is to consider India as a vital reference model and as the most preferable scenario for technology decisions and its viable forecast, below are a few points that are considered important in this paper:

1. Machine learning models can be a prime prediction and assessment tool in automated decision-making.
2. Accessing technology trends from patent data study and its forecast, using advanced explainable artificial intelligence (XAI). This will help in ensuring the technology's evolution tendency, as well as omissions that have created a research gap.
3. Carrying ahead the machine learning-driven predictions using open source platforms ca provide a vital research evolution statistics for long- and short-term policy decision-making.
4. In this paper, the author has brought ahead various aspects of data-driven decision-making and its inclusive efficiency in recreating silent zones of research domains. These will be the areas that need prior attention, during any science policy design.
5. The major objective behind this study is to track the technology efficacy of developed ideas and their exploration to establish a technology strategy. It includes wide range of applicability and forecast for technology management and process development. Objective information is a critical forecasting factor, and traditional decision-making makes it difficult to forecast objective results. However, in this paper, we have shown how a data-driven approach along with its systematic attempts can drive the way through a detailed development process at the macro level.

**Data-Driven Decision Making**

Data-driven decision-making is gaining popularity, in unveiling the essential objectives of black box AI outcomes. Logical decision-making and problem-solving, assessment have already been performing well, in generating policy scenarios and forecasting emerging trends (Bonvillian & van Atta, 2011). Perhaps, most of the scientific approaches have been studying technology emergingness under curve fitting & and stochastic models i.e. probabilistic distributions of potential outcomes, for technology predictions. On the other side, use of data-driven decision-making is raised by various decision-making bodies, government and intergovernmental organizations to share policy inputs. Surveillance capitalism with ethical design in creating networks is one of the useful approaches designed through this innovation approach and is used by several organizations to test and enlarge their digital ecosystems (Saura et al., 2021).

In this study it was observed that the concept of D3M or data-driven decision making was devised by **DARPA-Defence Advanced Research Project Agency**, which came into effective experimentation in 2016. It was the time when DARPA decided to create breakthrough technological developments and designs for military and civil defense, using advanced information technology tools. Through these non-data scientists but specific subject experts, who do not have data science as their expertise, are able to deal with automation data. D3M or data-driven decision making was incepted to automate the core task of machine learning

(Mattmann et al., 2018). Under this approach, many new models were developed like Alpha D3M, Two Raven, and MARVIN which were studied under the D3M augmentation to employ automated model discovery. Rather, D3M model discovery aims to enable users of subject matter expertise, to create requisite empirical models. This helps them to ascertain the real-world complex and empirical problem statements, as well as to enlarge the anchors of ICT and AI.

Through this paper, it was analysed that research and its forecast pattern have given evidence about data-driven information retrieval and assessment. This will help in reducing the gap between technology development and its integration for the elaboration of research landscape. Accordingly, many challenges also remain intact where and when the meaningful examination along with value-added trend can be extended. Here, many non-coordinated affairs are withstanding like data operational charges, network latency, insufficient computing clusters, and node classification to deal with this tranquillity. The emerged user-generated data and other relevant information includes the evidence in the form of a digital footprint corpus (Saura et al., 2021).

Industrial applications of automated valuation and prediction to settle for high-stake decisions will help in offering gains for sustainable process generation, using AI. Thus, DARPA defines this dawn of AI in different stacks as a wave for knowledge generation and knowledge systems, typically based on a specific set of rules designed through an expert system. Another lineage reflects statistical learning, which can be found today at various operational & and demonstration levels of unstructured data monitoring, and is not a part of the automation stratum.

To bring this subject ahead in upfront technologies, DARPA has started the explainable artificial intelligence (XAI) initiative which is based on the concept of white box and AI interpretability. Increased usage of AI systems therefore, has created a new start for research in explainable systems (Gunning & Aha, 2019),with its influential applicability results, bounded in observation of transparency, speed & and accuracy. The advent of data accessibility and analysis techniques has provided ample opportunities for research in almost all integrated areas placing its interest on tracking technology trends (Sreedharan et al., 2021). The first wave of AI indicates a lower index of perceiving information, and abstracting efficiency but had an eligible amount of reasoning capacity, for developing technology era at the new upfront.

The second wave therefore has a well-defined index of perceiving information and learning to expand it as models of classification and prediction. This form of data learning, which is done by separating manifolds of data structure, creates the new dimensional learning until the data manifolds are singly separated into individual units. Neural nets here, play a key impressive role, which learns from data by computing outputs from the received inputs, and then by adjusting its weight by error propagation (Gunning & Aha, 2019). A generalized neural network architecture is presented herein (fig.1 and 2)

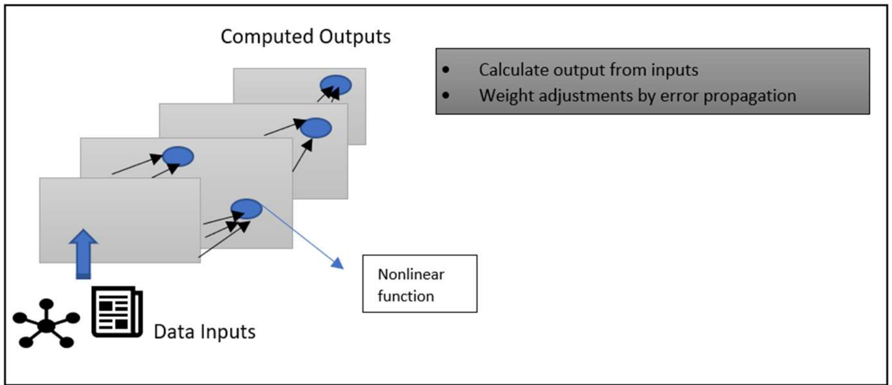



Image showing how neural nets learn from data

**Fig.1**

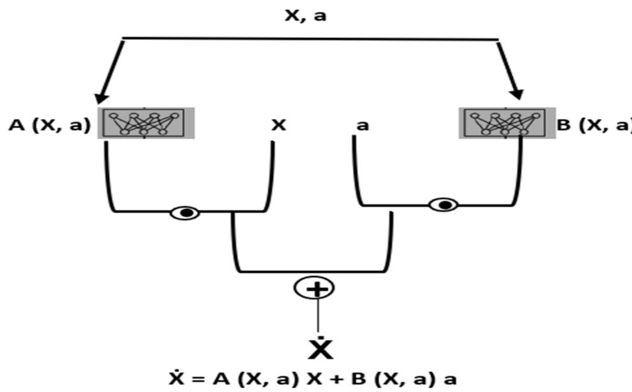



**Fig.2 Model Architecture for Predictive Control**

In (figure 2) structured neural architecture for algorithm-based predictive control is explained. It shows the final layer of **A** subnet of the same dimension as the state space & final layer of the **B** subnet must be of the same dimension as the control space. The network then computes a function of the form **($\dot{X}$ = A.X + B.X).** This structure makes it easier to recover time-varying derivatives (i.e., **A&B**) to use in predictive control algorithms. From the architecture of neural networks defined for predictive control, it was identified that model-free policy methods require an extensive amount of data with its statistical analysis, making it difficult to evaluate real-world variable/data assessment. On the other hand, model-based policy methods are more data efficient and often offer an easier path to apply in a real-world scenario (Broad A., Abraham,2019). The smartness of data estimation and its further application depends on the effectiveness of explainability, through relevant & and cognitive AI modules in terms of understanding and interpretation.

## Explainable Artificial Intelligence - Explainable Interpretation

With the dawn of more rationalized and convincing algorithms via data-enabled machine learning methodology, the adoption of XAI has been put forth immensely for industrial and large scale operations. Deep learning of datasets, at its prime performance, allows the machines to learn and automate via the discover and learn process, signifying the hierarchical data presentation. Here, data representation is executed by using a large dataset to train and test complex associations for predictiveness, but have their weak ability to describe uncleared networking mechanisms. Such, mechanically inferred models with improvised efficacy, are referred to as black box models. Black box models are here to display features of deep learning with uncleared predictions (Linardatos et al., 2020). To interpret the non-interpretable black box model, this approach created by DARPA captures the breadth and scope of Artificial intelligence and machine learning. It serves the purpose of the ability to explain their rationale, characterize the strengths and weaknesses of the model and its unbiased interpretation, and future performance of the model-generated system.

Through the study of this paper, it was identified that explainable artificial intelligence (XAI) is a less studied and less researched domain, where most of the scientific community's attention has been directed towards predictive analytics as well as methods of prediction, instead of going behind prediction processes and it's domain-specific multidisciplinary applications.

European Parliament 2018 has adopted **GDPR**-*General Data Interpretation Regulation* as law towards automated decision-making and profiling. It can be traced as a supplier of meaningful explanations of logic involved, in cases of automated information assessment. It has become a directive need during today's research, towards the dire need for meaningful explanations from trained models. These trained models originated out of cultured datasets, to fetch information under regression and classification statistics.

Nowadays, AI and deep learning of available data, are taking a fast pace in generating algorithms for text classification, pattern recognition, and speech recognition at the stride of human competitive results. It renders a par excellence, along with black box predictor interpretation. A hybrid solution is often required in the present time to solve problems, where the algorithms are combined to provide a concrete solution and a visible interface of best-fit algorithms (Guidotti et al., 2019). Conversion of black box models into transparent and interpretable models has now become a field of productive research while formulating automated decisions being used widely by researchers, policymakers, and analysts.

Fundamental questions about the rationale of the decision-making process at humans as well as a level of automation, which refers to the class of systems that have an expandable insight into how a decision is put ahead and how a model will behave are some invincible features of *Explainable Artificial Intelligence (XAI)* (Guidotti et al., 2019).

Interpretability, explainability & and transparency are a few of the major characteristics that explain system essentiality in research & and policy making, which uses data operability. However, many machine learning models are inherently available with explainability, available in post hoc and intrinsic transparent models of machine learning. For such ML models, post-hoc and model-specific XAI methods can be adopted for feature-relevant explanation, architecture modification, and visual explanations (Yang et al., 2022). Here it was analysed that more complex models provide less explainability due to the nature of the problem and the channel it follows. The less complex models may still perform with better explanations. Originally, mathematical explanations were available for AI systems that use inherent techniques like linear regression, decision trees, etc. In other words, a model of machine learning must reflect the reason codes, feature explanations, error identification for the prediction and interpretations generated from the networking of the black box.

Categorically, it shares local and global model behaviour, where the former explains the all-inclusive model behaviour and the latter explains each instance behaviour. These explanations offer tremendous benefits depending upon the machine's ability to explain the outcome taken, and its further utilization to understand the data and source. Explainable AI helps users to understand, trust, and manage the incoming generation of artificially intelligent machines, using logical inferences. Local interpretations are more particular and easier to follow, instead of global explanations. Such platforms are created by associating explanatory semantic information with features of learning simpler models designed as or having the capability of successive feature learning and its relative associations (Gunning & Aha, 2019). The author states that XAI is needed to enter logical pursuits of peeking into the black box model of machine learning interpretability. According to (Linardatos et al.2020b), a more interpretable machine learning system developed in and around the decade makes it easier to identify cause and effect relationships of results generated. European Union in this case has critically given its attentive mandate to create it as a research avenue of urgent technical focus for public policy. Recognizing its evitability for research and development it serves instrumental importance to be used in science policy areas (Ridley, 2022).

*Patent Mapping and Applicability of Explainable Artificial Intelligence Toward Technology Valuation*

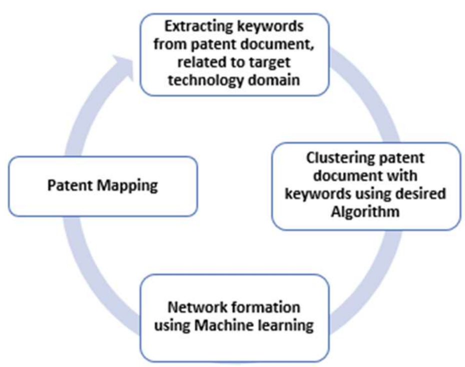
With the evolution of technology at a faster stride, technological forecasting is also taking an acceptable shape by most technology experts. Earlier forecasting results may signal trends, which can assist in preparing economic scenario-based healthy competition at the industrial development & and research level (Byungun Yoon, 2012). Technological forecasting is unavoidable when it will be discussed on a path for innovative index and innovative economy. During the study of innovativeness, patents are the primary output of R&D measurement of any firm's research scale. Utilization of patent data in creating patent maps, to cite important technological and innovation information is not limited merely to protecting legal rights for innovatively developed technology, rather it is also helpful in technology management. Hence, they represent the origin and features of newly developed technology and as an indicator of techno-economic development. Patent data maps can empirically be utilized for conceptual or quantitative analysis of technological changes (Choi & Park, 2009). Cumulative patent counts reflect the technology life cycle which in turn can be used to analyze the streamlined technology development stage (Trappey et al., 2011), to generate forecasting decisions based on forecast trend maps, visualization & and prediction (automated and statistical). The use of patent forecasts will support researchers and policymakers in distinguishing technology progress and expansion. This affirmation of technology expansion using well-supported database information may continue resource investment. It will also reflect, how and where to switch research directions, as an important perspective for every policy roadmap design (Trappey et al., 2011).

**Technology Success and Patent Economy**

As discussed above, patents chief participant in identifying technology outputs of the economy and its afterward economic suspension. Technology transfers are important visible innovations to examine R&D forecasts, with market strategies creating a valuable vault for intangible assets. Many countries like the United States, the European Union, Japan, and China have numerous programs for the protection of intellectual property rights that are necessary for technology transfer ideas and to increase the proportionality of innovation (Fredström et al., 2021). Although the number of technology transfers is lower than several patents applied and granted, patentability norms are a must to consider in extrapolating the **g**lobal **s**cientific **e**conomy. With everyday increasing demand for technology and its demanding stalwarts for society, secured information processing has also become a significant factor to consider it precisely. This has led the way to move ahead in the direction of superior and systematic strategic investment and development in innovation areas.

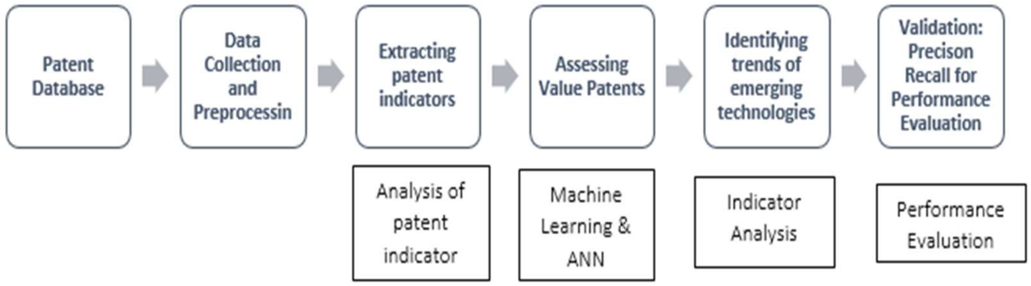
Our paper, (Fig.3 &4) explains the methodology of patent management and its resource data analysis using AI interpretation.

Considering the above discussion, patent analysis is a useful method in innovation forecasting patterns to help them participate and establish developmental technologies for the research avenues (Cho, H.P. et. al,2017).




**Process of Patent Mapping**

**(Fig.3)**




**(Fig.4) Process of Proposed Methodology Adopted and Patent Mapping for analysis and role of Explainable Artificial Intelligence**

The above figure shows a methodology that has been undertaken in this paper to reflect, how patent data can help in the analysis of globalized technology forecasting. Given the multi-facet approach to identifying emerging technologies and their results in modeling, most of such practices are majorly expert-centric like Delphi and other large-scale survey methods. Although major expert behavior tends to land at best scientific decisions, at the same time, it was identified that expert-centric approaches are excessively time-consuming. Hence, the path taken to attain these decision scenarios may lead to some essential bias and give rise to a fallacy in decisions at all verticals. Technology assessment is consecutively a driver for technology scale-up & and forecast by finding how patents can prove beneficial towards industrial development. It is another name for a class of policy studies that attempts to address the wide group of societal impacts of newly introduced technologies (Tran & Daim, 2008). A few other methods are also available, which were used for assessment, as a standard approach in forecasting assessment including statistical & and empirical practices. Methods, which are followed by statisticians as well as policy analysts are here below and are the most prominent methods of forecasting used to date (Firat, Woon et.al. 2008)

- Expert Opinion
- Trend Analysis
- Monitoring and Intelligence Methods
- Statistical Methods
- Modelling and Simulation
- Scenario analysis
- Economic modelling methods
- Descriptive assessment using heterogeneous analysis.

Conventionally accepted forecasting methods help in understanding technology valuation, under the machine learning approach, adopted widely to consider, classify, forecast & and elaborate decisions on the best-fit features available in the database. Various arguments have shown that machine learning models show high accuracy and high prediction rates, as compared to conventional prediction methods, and can be performed using both econometric or empirical methods and machine learning models. Till now econometric models analyze forecasting using exponential smoothening models, regression models, and ARIMA (Auto-Regressive Integrated Moving Average) models, which have been used continually by researchers. However, under limitations of the econometric forecast, these methods can represent the possibility of overfitting and forecasting data bias and its error rates thus generated.

The author of this paper has used Explainable Artificial Intelligence (XAI) as one of the leading efforts under ML research challenges as a method to interpret machine learning and assessment of data inputs. It promotes a set of techniques, algorithms, and explanations for methodology adopted for data interpretation for human-understandable descriptions. It has been identified through varied research that the said domain is proving beneficial and an upcoming area of advanced AI, as proven by (Gunning & Aha, 2019), (Linardatos et al., 2020), (Sreedharan et al., 2021), (Dazeley, Richard, et al., 2021). Consequently, changes for the rising of XAI in systems have been approved legislatively by the European Union's new General Data Protection Regulation. Therefore, it started coming into force, with the rise of major open-access programs like bons.ai of Microsoft, H2O.ai, Amazon AWS, etc.

Similarly, one of the explainable AI platforms that came across the forecasters and has also been used in this paper for patent investigation is **H2O.ai**. It provides a sufficient resource to generate essential data results. H2O.ai is an open-source automated machine learning platform designed to scale large datasets using R, Scala, Python, and even Java-based APIs as easily deployable solutions. It is a highly scalable and fully automated platform with supervised and unsupervised learning to train the data from a large section of individual models within a single function (Linardatos et al., 2020b).

H2O.ai provides its experimental run characteristics in the form of a leaderboard having a list of models it has followed and characterized visualization to get a fair idea of individual database features. It trains a variety of algorithms like Gradient Boosting machine, XGBoost, Random Forest, Deep Neural Networks, etc. making it a user-friendly machine learning software with simplified yet more explainable features, easier to understand and predict. Patent data contains a variety of information such as publication and application date, along with several global patent applicant statuses. Other researchers use natural language processing to reflect the type of technology explained and data evaluation, in the patent database. Methods used by the analyst may range from text mining to data and natural language processing (Krestel et al., 2021).

## Methodology

In this paper, a patent database prediction approach has been taken to identify emerging technology modeling for the available essential variables. These analytical AutoML methods will help in the corresponding analysis of expert judgments and decisions. For, this paper, we have taken data OECD database to provide evidence for the patent's value to identify of emergingness from noted observations. Quantitative indicators then further infer technology emerging from the given patent database.

In this paper, the OECD (Organisation for Economic Cooperation and Development) dataset has been taken as a reference which provides patent inputs from around the world. The data covers the following characteristics:

- Patent applications to (EPO)European Patent Office from 1999 onwards.
- Patent granted by EPO.
- Patent applications to (USPTO) US Patent & Trademark Office
- Patents granted by USPTO.
- Patents filed under (PCT) Patent Co-operation Treaty, which is an international phase that designates EPO.
- Patents that belong to Triadic Patent Families as per OECD definition, which are the subset patents filed together at EPO, JPO-Japan Patent Office, and USPTO, to protect the same set of inventions.
- Patents that belong to IP5 families i.e. those patents protected in at least two IP offices worldwide. These patents are examples that have been filed in five IP offices (IP-5) namely EPO, JPO, and USPTO and Korean Intellectual Property Office (KIPO) and the People's Republic of China National Intellectual Property Administration (CNIPA).

**Dataset Detail**

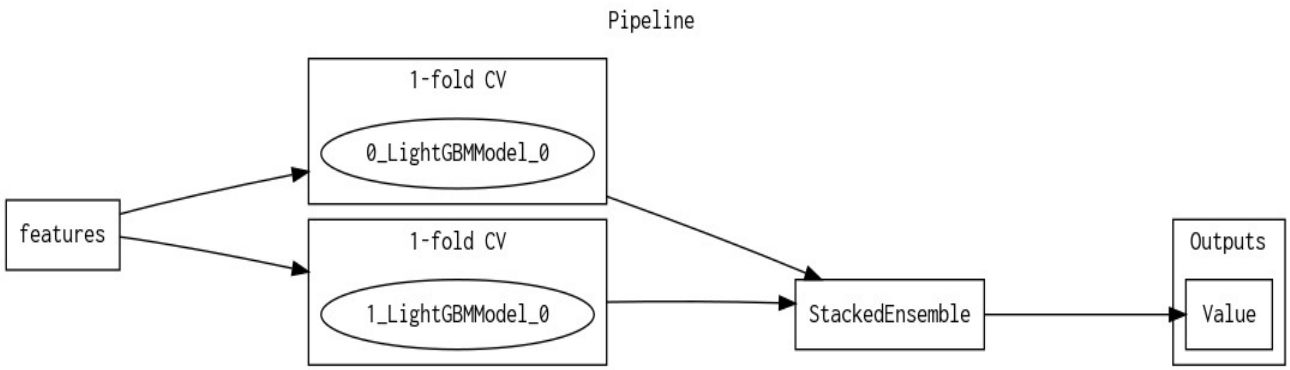
A total of 153790 cells are present in the database and are then re-construed according to the type of information (Patent number, assignee, class, and year). In addition to this, the patents which are published by the assignees are also represented as inventors and applicants. This patent database thus has information about the assignee patent as applicant and inventor, along with the year of the patent granted. The patents office and patent family indicator in the database indicate the type of patent family, in which the assignee as inventor or applicant has chosen to register its patent in any specific geographical division, ruled by the patent rules. The patent citation index is not considered in a database for identification of emergingness, as it should be noted that the time lag between citing patents and cited patents, took 3-5 years on average. The latest patents in an obvious manner take more time for citation and will have lower citation scores. This factor will influence the affinity of emerging news and tracing ahead of the latest effect on technology in the economy.

This study employs a total of five indicators under consideration to capture the key characteristics of emerging technologies:

1. Patents Office & Patents Families
2. Country
3. Technology domains

4. Time
5. Value of Patents

As the dataset is of medium dimensionality, a regression model is used to identify characteristics of datasets including the class of technologies that effectively represented a higher number of patents concerning time. This side-by-side helps in generating the overall feature importance generated out of the automated machine learning interpretation model.

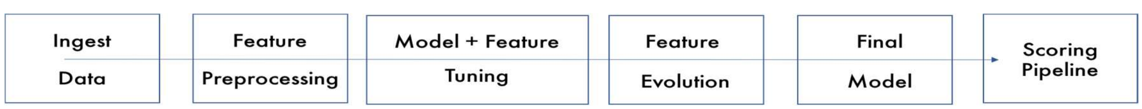



**Final Model Pipeline (Fig.5.)**

**Tree Interpretation: The fitted features of the final model are the best features found during the feature engineering iterations. The target transformer indicates the type of transformation applied to the target column.**

**Model**

For this experiment, the below-mentioned pipeline is used by driverless AI, to finalize the model and model parameters.




(Fig.6.)- Model Pipeline

To generate this model parameter, H2O.ai auto ML finds the best set of model parameters using feature transformation to finalize the model. This ML model performs supervised learning with a light regressor. The training dataset consists of both numerical and categorical columns to be used for the data ingestion stage for feature pre-processing.

For each dataset, the following process was carried out in XGBoost and LGBM: (i). The optimum parameters for each method were stratified using 3-fold cross-validation within the dataset using grid search, which compares the CV score for each combination of parameters and returns the set having the best parameter ruling. (ii) Best set of parameters dragged from grid search was used to train the corresponding ensemble using the whole set of data used;(iii) Additionally, the ensemble was trained for LGBM for each possible parameter combination. The table:1 shows the score and training time of LIGHTGBM models evaluated by Driverless AI. This stage combines random hyperparameter tuning with feature selection and generation. Features in each iteration are updated using variable importance from the previous iteration as a probabilistic before deciding what new features to create. The best-performing model and features are then passed to the feature evolution stage.

**Performance of Final Model obtained via H2o.ai iteration (Table.1)**

| Scorer | Final ensemble scores on validation | Final test scores | Final test standard deviation |
|---|---|---|---|
| MAE | 17.72352 | 18.65979 | 0.4869117 |
| GINI | 0.9657724 | 0.9627199 | 0.001552367 |
| MAPE | 698.9026 | 551.0227 | 253.9525 |
| MER | 36.68272 | 37.82015 | 0.4734313 |
| MSE | 2425.903 | 2918.89 | 373.9335 |
| R2COD | 0.9220475 | 0.9030651 | 0.01229241 |
| R2 | 0.9243182 | 0.9082405 | 0.009089597 |
| RMSE | 49.25345 | 54.02675 | 3.333005 |
| RMSLE | 1e+36 | 1e+36 | 0 |
| RMSPE | 49975.86 | 30951.3 | 15293.62 |
| SMAPE | 80.56298 | 81.45068 | 0.7178019 |

## Model Scoring

Scoring of auto ML models was performed using MAE, which is the Mean Absolute error, and it defines the average of absolute errors. MAE units are the same as the predicted target and are useful in understanding whether the size of the error is of concern in the model or not.

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}.$$

Or MAE = 1/n *∑ $|y_i - x_i|$

arithmetic average of absolute errors $|e_i| = |y_i - x_i|$

$y_i$ = prediction
$x_i$ = true values.

The fitted features of the final model are the best features found during feature engineering iterations, which have been shown in the feature correlation map or the feature importance shown in (fig:7).

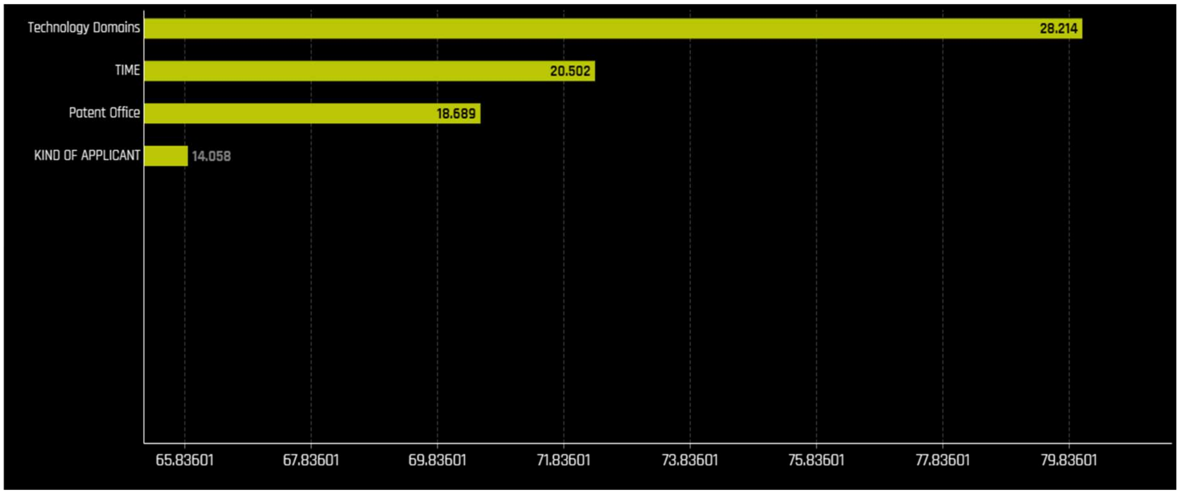
In this paper, we have considered MAE as the regression metric for model evaluation, as MAE takes all the individual differences equally in average.

Performance of Final Model-Model Scoring and accuracy index (Table 1). According to the model performance indicators we have followed the MAE scoring, for the model validation, as it provides the mean absolute difference between the predicted values and the actual value of the dataset. The lower the MAE, the better the model, to fit in the dataset.

## Model Interpretation and Patent Scenario

ML model of obtained data shows that the technology domain is playing an important role in identifying the featured importance of patent information obtained from OECD. Also, it shows that data obtained from OECD shows its subjective relative importance for various technology domains that must be assessed based on the value of patent scored in the model.
Classification of technology domains that have taken the lead in the last few years since 2009 are Performing Operations; Transporting, environment, ICT, Physics, Computer technology, Electricity, Chemistry; Metallurgy, Human Necessities, Pharmaceuticals, Organic fine chemistry, and Digital communication. These top subject patent performers come in sight with top predictors in patenting offices via offices of USPTO and PCT which are patent families that improve comparability and quality of patent indicators. Identifying these characteristics from patent data using the auto ML technique is a supportive asset for explainable AI and gives an in-depth analysis of indicators that are not followed during technology development compliances as well as policy formulation strategies. As mentioned in (Fig.7) is the partial dependency plot Partial dependence plots show the partial dependence as a function of specific

values for a feature subset. The plots under (Fig.7) show how machine-learned response functions change based on the values of an input feature of interest while considering nonlinearity and averaging out the effects of all other input features. Partial dependence plots enable increased transparency in a model and enable the ability to validate and debug a model by comparing a feature's average predictions across its domain to known standards and reasonable expectations.




**(Fig.7) The partial dependence plots are shown for the top 4 original variables. The top 4 original variables are chosen based on their Component Based Variable Importance.**

Global feature importance is a measure of the contribution of an input variable to the overall predictions of the random forest surrogate model. Global feature importance is calculated by aggregating the improvement in splitting criterion caused by a single variable across all the decision trees in the random forest surrogate model. Local feature importance is a measure of the contribution of an input variable to a single prediction of the random forest surrogate model. Local feature importance is calculated by removing the contribution of a variable from every decision tree in the random forest surrogate model and measuring the difference between the prediction with and without the variable. Both global and local variable importance is scaled reflecting the largest contributor importance in the dataset.

Results obtained have shown that patents are key measure statistics for innovation output as patent indicators, which reflect the inventive performance of countries, regions, technology domains, and registration firms. In this case, the data has been taken for India from 1990-2019-2020, for various technology domains of approximately 48 in numbers, also the number of inventors and applicants has been counted. The database obtained from OECD represents the country profiles in a global context with all the best possible data values obtained through it.

Accordingly, these indicators can be defined and calculated over the years depending on the requirement of time forecasting. Given the duration of a patent that has been mentioned in the data, interpretation for technology development and transformation can be defined as short--, mid-, and long-term technology planning.

The derived model has presented us the shapely explanation which presents local as well as global feature contributions along with bias term sum. Where in the case of classification problems, they sum to the prediction. This Naïve Shapely explanation given below calculates local shapely explanation for original features, which provides insights into feature importance which again comes out to be the technology domain. It reflects that the patent office and the time series mentioned in a data frame, cannot give the accuracy to identify technology transformation statistics, as well as the values of patents, are not independent of the technology domain. Hence the technology emergingness related to various technology domains reflects the patent data scenario, diving the innovativeness as well as the increase and decrease in technology sectors (Smith, 2005).

In this experiment, we have identified that patent data are of reflective importance in driving a country's technology eruption because:

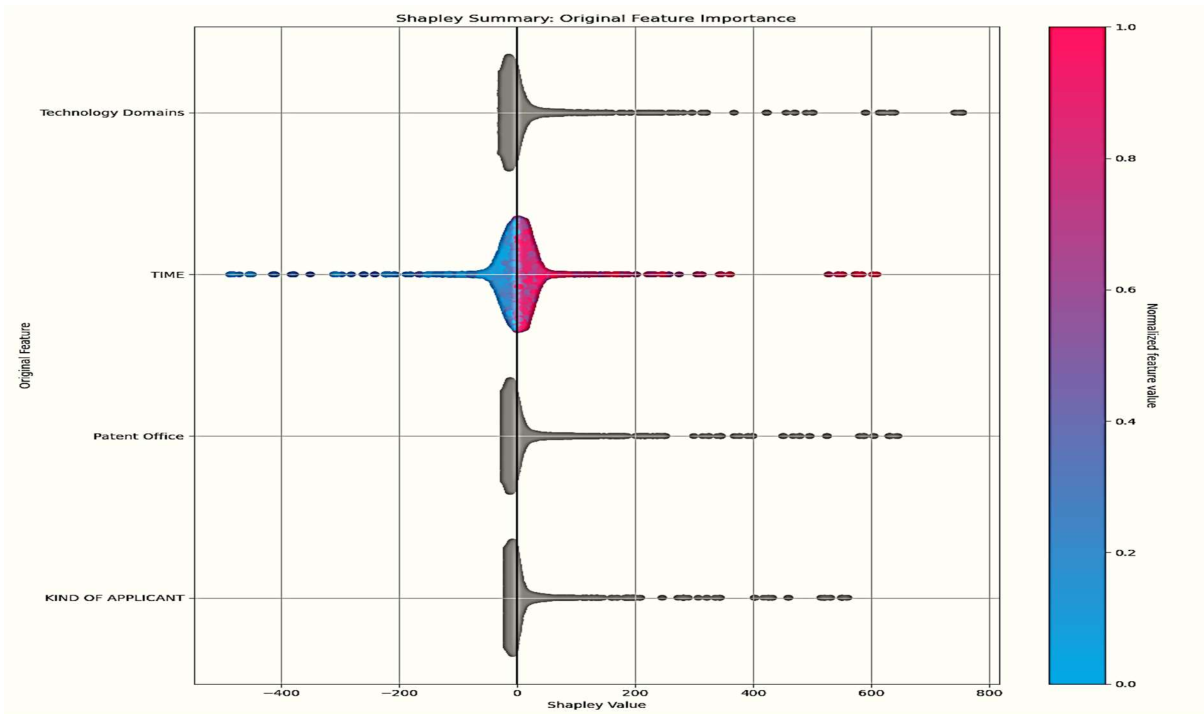
- Patents show closer acquaintance with any country's global inventiveness.
- Patent covers a broad range of technology domains, where other data sources cannot generate such insightful details, and hence the content of patent data is a rich source of information. They also include inventor details, country of origin, and applicant details.
- The data shows that patenting scenarios provide a more reliable shift and R&D focus.

## Results & Discussion

### Model Description & Inference

- XGBOOST and Light GBM Regression algorithm was performed in H2o.ai auto ML. Light GBM is an open-source framework, originally proposed by Microsoft®. It is a decision tree algorithm, which divides input layer parameters into different parts and then builds up mapping association between input and output. LGBM works on leaf-wise growth strategy, i.e. Gradient Boosting Decision Tree (GBDT)algorithm to reach more accurate predictions (Gan et al., 2021). Here the target output of each tree model's errors of the previous tree keeping the same input values. The GBDT algorithm generates final predictions by using ensemble estimation which will be closer to the output generated during tree formation. During, each vertical decision tree formation, the parameters that determine the structure, influencing the iteration process are determined through cross-validation. LGBM stands as a competitor for XGBoost in terms of both training time and accuracy.
- XGBoost is a decision tree ensemble-based gradient boosting algorithm, which is designed to be highly scalable as it works on level-wise tree generation and creates the node up to the max_depth specified. It builds an additive expansion of the objective function by minimizing the loss function. Decision trees are the main classifiers in XGBoost and thus variation in the loss function, which controls the tree complexity by faster training and requires less storage space. XGBoost implements numerous methods to increase the

training speed of decision trees not directly related to ensemble accuracy. It focuses on reducing computational complexity for finding the best split, by enumerating the best candidate split and selecting the one with the highest gain. This function is done by performing a linear scan over each sorted attribute, to find the best split for each node using multiple additive regression trees.




(Fig.8) Shapely summary plot of Feature Importance

## Conclusion

### In the provision of Implementation of our approach

H2o.ai methodology used in the form of advanced auto ML and driverless AI to employ explainability, which is an open-source platform, should be carefully deployed in practice, as it further influences the decision-making approach.

The proposed methodology can be advanced and replaced using many other machine learning models as well as the feeding data and its type. As mentioned in (table: 5) most of the scoring process shows sufficiently better performance, but in this experiment, we have taken MAE as the major scoring for the proposed model.

Our method can serve as a monitoring method to identify emerging technologies, as the data and technology used to obtain results can produce reproducible results. If new patents are issued, extraction of input indicators, rise in the patents of emerging technologies, and innovation dependencies can be predicted and plotted using only certain additional model requirements. Finally, the model developed should be reviewed and updated, to measure more precise technological scenarios.

**Scenario Derived for Technology Policy**

As the development of new and innovative technology policy requires robust engineering of technology outcomes and innovation as well as technology resilient indicators, this plays an important pursuit. Technology-based products create business value in today's economy and such advanced forecasting methods can create a crucial demanding platform. Accordingly, a novel systematic methodology has been proposed in this paper using auto ML, which can be easier to use and adopted easily by non-data scientists/ subject matter experts, It is based on predictions of patent growth, its diversity, and technology evolution within the given time.

As per the DARPA landscape of developing explainable artificial intelligence (XAI) for future endeavors, this proposed methodology acts as a way forward for transparent interpretability yet has many limitations. In this study, it has been identified that algorithmic explainability and its inscription for the decided revealed information system, prove a truly important facet where many black box models are on the desk to propose inbound data communications.

Our research finds that considerable research in policy and decision-making can be triggered at the right pace, by using such XAI and white box models, which can propose open-access results to every subject matter expert. However, evaluating these methods running with larger datasets can still pose a greater challenge to the research community & and in front of XAI users. User studies have identified that typical graphical representations can not only be used to fetch data results but such proposed methods should be used in identifying real-time decision-making. Recent developments in human-ML-XAI-guided explanations to evaluate decision-making show promising improvements in technological landscapes.

**Acknowledgments**

We are thankful to the Visvesvaraya scheme of the Ministry of Electronics, Information Technology (MEITy), for their support in carrying out this research of capability building and policy experimentation.

**Funding**

- Partial financial support was received from the Visvesvaraya scheme of the Ministry of Electronics, Information Technology (MEITy)
- Grant Agreement No: [**MEITY-PHD-3258**]

**Author Contribution:**

- **1st Author**: The first author, here is a research scholar who is pursuing her studies in technology policy and its implementation in the science and technology policy domain. The author is also passionate about pursuing policy research and its divergent areas to be taken care of.

- **2nd Author**: The second author is a senior and eminent Professor at Jamia Hamdard Institute and has diverse and wide experiences in the science policy domain.
- **3rdAuthor**: The third Author is a senior Professor at Jamia Hamdard Institute and has diverse and wide experiences in science policy and IT domains.
- **4thAuthor**: The fourth author is a Professor Jamia Hamdard Institute and has diverse and wide experiences in science policy and IT domain.
- **5thAuthor**: The Fifth Author is a senior Professor at Jamia Hamdard Institute and has diverse and wide experiences in the science policy domain.
  - ✓ **1stAuthor**: Zeba Hasan is the primary author and has written the whole manuscript, including work on the ARIMA model algorithm.
  - ✓ **2ndAuthor**: Prof. M. Afshar Alam - reviewed the manuscript.
  - ✓ **3rd Author**: Dr. Harleen Kaur - reviewed the manuscript.
  - ✓ **4th Author**: Dr Bhavya Alankar research co-supervisor
  - ✓ **5th Author**: Ihtiram Raza Khan research co-supervisors

**Conflict of Interest:**

- Author *Zeba* received only partial support from the Visvesvaraya scheme of the Ministry of Electronics, Information Technology (MEITy*)* and is a student of Prof. M. Afshar Alam, Dr. Harleen Kaur, Dr Bhavya Alankar, Ihtiram Raza Khan
- No, I declare that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

**Data Availability Statement**

The data taken into consideration in this manuscript is open-source data and can be made available.

## References


1. Bonvillian, W. B., & van Atta, R. (2011). ARPA-E and DARPA: Applying the DARPA model to energy innovation. The Journal of Technology Transfer, 36(5), 469–513. https://doi.org/10.1007/s10961-011-9223-x
2. Broad, A., Abraham, I., Murphey, T.D., & Argall, B. (2018). Structured Neural Network Dynamics for Model-based Control. ArXiv, abs/1808.01184.
3. Byungun.Yoon,M.(2012,January)., Articles–Manupatrahttps://Articles.Manupatra.Com/#.https://articles.manupatra.com/article-details/Applicability-of-Patent-Information-in-Technological-Forecasting-A-Sector-Specific-Approach.
4. Cho, H.P., Technological Forecasting & Social Change (2017), https://doi.org/10.1016/j.techfore.2017.11.012
5. Choi, C., & Park, Y. (2009). Monitoring the organic structure of technology based on the patent development paths. Technological Forecasting and Social Change, 76(6), 754–768. https://doi.org/10.1016/j.techfore.2008.10.007

6. Choi, Y., Park, S., & Lee, S. (2021). Identifying emerging technologies to envision a future innovation ecosystem: A machine learning approach to patent data. Scientometrics. https://doi.org/10.1007/s11192-021-04001-1
7. Dazeley, Richard & Vamplew, Peter & Cruz, Francisco. (2021). Explainable Reinforcement Learning for Broad-XAI: A Conceptual Framework and Survey
8. Faloutsos, C., Gasthaus, J., Januschowski, T., & Wang, Y. (2019). Classical and Contemporary Approaches to Big Time Series Forecasting. Proceedings of the 2019 International Conference on Management of Data. https://doi.org/10.1145/3299869.3314033
9. Firat, Woon, Madnick, A. K. W. L. S. & Composite Information Systems Laboratory (CISL) Sloan School of Management, Room E53-320 Massachusetts Institute of Technology Cambridge, MA 02142. (2008, July). Technological Forecasting – A Review. https://skat.ihmc.us/rid=1NMM9Z943-1197XJF-5CGB/2008-15.pdf
10. Fredström, A., Wincent, J., Sjödin, D., Oghazi, P., & Parida, V. (2021). Tracking innovation diffusion: AI analysis of large-scale patent data towards an agenda for further research. Technological Forecasting and Social Change, 165, 120524. https://doi.org/10.1016/j.techfore.2020.120524
11. Gan, M., Pan, S., Chen, Y., Cheng, C., Pan, H., & Zhu, X. (2021, May 3). Application of the Machine Learning LightGBM Model to the Prediction of the Water Levels of the Lower Columbia River. *Journal of Marine Science and Engineering*, *9*(5), 496. https://doi.org/10.3390/jmse9050496
12. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. ACM Computing Surveys, 51(5), 1–42. https://doi.org/10.1145/3236009
13. Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. AI Magazine, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850
14. Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. AI Magazine, 40(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850
15. Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy, 23(1), 18. https://doi.org/10.3390/e23010018
16. Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020b). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy, 23(1), 18. https://doi.org/10.3390/e23010018
17. Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020b). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy, 23(1), 18. https://doi.org/10.3390/e23010018Krestel, R., Chikkamath, R., Hewel, C., & Risch, J. (2021). A survey on deep learning for patent analysis. World Patent Information, 65, 102035. https://doi.org/10.1016/j.wpi.2021.102035

18. Martino, J. P. (2003). A review of selected recent advances in technological forecasting. Technological Forecasting and Social Change, 70(8), 719–733. https://doi.org/10.1016/s0040-1625(02)00375-x

19. Mattmann, C., Shah, S., & Wilson, B. (2018). MARVIN: An Open Machine Learning Corpus and Environment for Automated Machine Learning Primitive Annotation and Execution. ArXiv, abs/1808.03753.

20. OECD. (n.d.). OECDPatentStatisticsManual.https://www.oecd.org/sti/inno/oecdpatentstatisticsmanual.html

21. Personal Experiences and relative framing ideas to design better policy approaches.

22. Ridley, M. (2022, June 15). Explainable Artificial Intelligence (XAI). *Information Technology and Libraries*, *41*(2). https://doi.org/10.6017/ital.v41i2.14683

23. Saade, Marie & Jneid, Maroun & Imad, Saleh. (2019). Predicting technology success based on patent data, using a wide and deep neural network and a recurrent neural network.

24. Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. 2013 International Conference on Collaboration Technologies and Systems (CTS). https://doi.org/10.1109/cts.2013.6567202.

25. Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. 2013 International Conference on Collaboration Technologies and Systems (CTS). https://doi.org/10.1109/cts.2013.6567202

26. Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. International Journal of Information Management, 60, 102331. https://doi.org/10.1016/j.ijinfomgt.2021.102331

27. Schoenborn, Jakob & Althoff, Klaus-Dieter. (2019). Recent Trends in XAI: A Broad Overview on Current Approaches, Methodologies, and Interactions.

28. Shobana, G., & Umamaheswari, K. (2021). Forecasting by Machine Learning Techniques and Econometrics: A Review. 2021 6th International Conference on Inventive-Computation-Technologies-(ICICT). https://doi.org/10.1109/icict50816.2021.9358514

29. Smith, G. W. (2005). Intellectual Property Rights, Developing Countries, and Trips. The Journal of World Intellectual Property, 2(6), 969–975. https://doi.org/10.1111/j.1747-1796.1999.tb00102.x

30. Sreedharan, S., Chakraborti, T., & Kambhampati, S. (2021). Foundations of explanations as model reconciliation. Artificial Intelligence, 301, 103558. https://doi.org/10.1016/j.artint.2021.103558

31. Tran, T. A., & Daim, T. (2008). A taxonomic review of methods and tools applied in technology assessment. Technological Forecasting and Social Change, 75(9), 1396–1405. https://doi.org/10.1016/j.techfore.2008.04.004

32. Trappey, C. V., Wu, H. Y., Taghaboni-Dutta, F., & Trappey, A. J. (2011). Using patent data for technology forecasting: China RFID patent analysis. Advanced Engineering Informatics, 25(1), 53–64. https://doi.org/10.1016/j.aei.2010.05.007.

33. Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black box for the medical explainable AI via multi-modal and multi-center data fusion: A mini-review, two showcases, and beyond. Information Fusion, 77, 29–52. https://doi.org/10.1016/j.inffus.2021.07.016