|  |  | **Paper No** | **Model/System** | **Method** | **Application** | **Description** | **Software** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Applied Machine Learning in Civil Engineering Informatics*** | **Computer-Aided Informatics** | 1 | Firefly algorithm (FA), metaheuristic (Meta) intelligence, and least squares support vector regression (LSSVR) (MetaFA-LSSVR) | **FA + LSSVR** | Currency exchange rate forecast | The MetaFA automatically tunes the hyperparameters of the LSSVR to construct an optimal sliding-window LSSVR prediction model | MATLAB |
| 2 | Stock prices prediction for nonlinear time series |
| 3 | Smart artificial firefly colony algorithm‐based support vector regression (SAFCA‐SVR) | **FA + chaotic maps + adaptive inertia weight + Lévy flight + LSSVR** | 1. Predicting concrete compressive strength
2. Predicting resilient modulus of subgrade soil
3. Estimating cooling load
 | Smart artificial firefly colony algorithm‐based support vector regression (SAFCA‐SVR) system that integrates firefly algorithm (FA), chaotic maps, adaptive inertia weight, Lévy flight, and least squares support vector regression (LS‐SVR) | MATLAB |
| **Application to Geotechnical Engineering** | 4 | Estimate the peak shear strength of Fiber-reinforced soil (FRS) |
| 5 | Geogrid‐inspired nanostructure | **Nano‐geogrid (denoted as NWD)** | Reinforce side slopes and retaining walls | The CZTS nanowall electrode reinforced by the nano‐geogrid (denoted as NWD) shows not only remarkable mechanical and electrochemical stability but also considerable electrochemical performances | - |
| 6 | Tiering SVM-(SVR/SVR) | **(1) Classification and regression methods, i.e., REG, CART, GENLIN, CHAID; (2) Machine learners, i.e., ANN, SVM, SVR; and (3) Metaensemble models, i.e., voting, bagging, stacking, and tiering** | Predicting the peak friction angle of fiber-reinforced soil (FRS) | The optimal model obtained after further model training, cross-validation, and testing was the Tiering SVM-(SVR/SVR) method | IBM SPSS Modeler |
| 7 | Smart firefly algorithm (SFA)] and machine learning [least squares support vector regression (LSSVR) | **(CART, GLR, and SVR), their ensembles (CART + GLR, CART + SVR, GLR + SVR, and CART + GLR + SVR), and (SFA-LSSVR)** | Predicting reinforcement tensile loads and its feasibility for facilitating early designs of geosynthetic-reinforced soil (GRS) structures | The analytical results demonstrate that the SFA-LSSVR is superior and outperforms the other models | MATLAB & IBM SPSS Modeler |
| 8 | Tiering SVM-(SVR/SVR) | **(1) Classification and regression methods, i.e., REG, CART, GENLIN, CHAID; (2) Machine learners, i.e., ANN, SVM, SVR; and (3) Metaensemble models, i.e., voting, bagging, stacking, and tiering** | Predicting the peak friction angle of fiber-reinforced soil (FRS) | The optimal model obtained after further model training, cross-validation, and testing was the Tiering SVM-(SVR/SVR) method | IBM SPSS Modeler |
| **Application to Structural Engineering** | 9 | Smart artificial firefly colony algorithm‐based support vector regression (SAFCA‐SVR) | **FA + chaotic maps + adaptive inertia weight + Lévy flight + LSSVR** | Shear Strength Prediction in Reinforced Concrete Deep Beams | Smart artificial firefly colony algorithm‐based support vector regression (SAFCA‐SVR) system that integrates firefly algorithm (FA), chaotic maps, adaptive inertia weight, Lévy flight, and least squares support vector regression (LS‐SVR) | MATLAB |
| 10 |
| 11 | GLM with K-means clustering | **This work constructed nine prediction models; three are directly generated by the GLM, simple regression, and ANNs The mixed model of K-means and two-step clustering generated three prediction models**  | Predicting the aseismic ability of school buildings | The prediction model that performs best is that built using the GLM with K-means clustering | ETABS & IBM SPSS Modeler |
| **Application to Construction Materials** | 12 | Smart firefly algorithm (SFA)] and machine learning [least squares support vector regression (LSSVR) | **Single and ensemble models (MLP, SMOreg, REPTree, and LR) + SFA-LSSVR** | Predicting the shear strength of RC beams | Analytical results verified that the hybrid AI model (SFA-LSSVR model) outperformed all of the other single and ensemble AI models | MATLAB & WEKA |
| 13 | Artificial neural network (ANN) model in association with a modified firefly algorithm (MFA) | **MFA-ANN & SFA-LSSVR** | Predicting compressive and tensile strength of high-performance concrete | The analytical results demonstrate that the MFA-ANN is superior and outperforms the other models | MATLAB  |
| 14 | [Smart firefly algorithm (SFA)] and machine learning [least squares support vector regression (LSSVR)] | **Ensemble models (ANNs, SVR/SVMs, CART, and LR) and (SFA-LSSVR)** | 1. Predicting the pitting corrosion risk
2. Predicting marine corrosion rate
 | SFA-LSSVR model had the lowest MAPE values for predicting corrosion rate and pitting risk | MATLAB |
| 15 | Ensemble models that integrate multiple classifiers, the voting, bagging, tiering, and stacking combination methods | **Individual ML models (MLP, SVM, CART, and LR) + ensemble models ( voting, bagging, and stacking)** | Predict the compressive strength of HPC | The stacking based ensemble model composed of MLP/CART, SVM, and LR as the first level models and SVM as the second level model performed best in terms of MAE, RMSE, and MAPE | WEKA |
| 16 | **Individual ML techniques (ANNs, CART, CHAID, LR, GENLIN, and SVMs) + Ensemble models** | All ensemble models achieved good prediction performance (R ⩾ 933%) and significantly improved the error rates from 42% in previous reports to 697% | IBM SPSS Modeler |
| 17 | Evolutionary fuzzy support vector machine inference model for time series data (EFSIMT) | **Fuzzy Logic (FL) + weighted Support Vector Machines (wSVM) + fast messy genetic algorithms (fmGA)** | Predict the compressive strength of HPC | Validation results show that the EFSIMT achieves higher performance in comparison with Support Vector Machines (SVM) and obtains results comparable with Back-Propagation Neural Network (BPN) | MATLAB |
| 18 | Hierarchical classification and regression (HCR) | **SVM is trained by different numbers of classes in the first level of HCR and in which LR, MLP, and SVR are the regression models in the second level of HCR** | Predict the compressive strength of HPC | HCR with a 4-class support vector machine in the first level combined with a single ANNs obtains the lowest mean absolute percentage error | WEKA |
| 19 | Multiple additive regression tree (MART) | **MART, BRT, SVM, ANNs, and MR** | Predict the compressive strength of HPC | The cross-validation of unbiased estimates of the prediction models for performance comparison purposes indicated that multiple additive regression tree (MART) was superior in prediction accuracy, training time, and aversion to overfitting | WEKA & IBM SPSS Modeler |
| **Application to Water Resource & Hydraulic Engineering** | 20 |  [Smart artificial firefly colony algorithm (SAFCA)] and machine learning [least squares support vector regression (LSSVR)] | **Individual ML models (MLP, SVM, CART, and LR) + The ensemble models (voting, bagging, and stacking) + Hybrid model** |  Predict decisive information in water quality management | ANNs model is recommended for making predictions of water quality when fast and simple data analytics is required | WEKA, WEKA, IBM SPSS Modeler, Rapid Miner Studio, Azure Machine Learning and MATLAB |
| 21 | [Smart artificial firefly colony algorithm (SAFCA)] and machine learning [least squares support vector regression (LSSVR)] | **Individual ML models (ANNs, CART, CHAID, and GENLIN) + Ensemble models + Hybrid models (SAFCA-LSSVR)** | Predicting bridge scour depth near piers | The proposed SAFCAS system provides an effective knowledge-based tool for making rapid and accurate predictions that can aid decision makers in solving practical problems | MATLAB & IBM SPSS Modeler |
| 22 | Genetic algorithm (GA)-based support vector regression (SVR) | **CART, CHAID, MLR, and ANN + GA-SVR** | Predict bridge scour depth near piers and abutments | The hybrid GA–SVR model provided error rates that were 813% to 964% more accurate than those obtained by other models when predicting scour depth | IBM SPSS Modeler & MATLAB  |
| ***Building Energy Management*** | **Residential Buildings** | 23 | SARIMA-PSO-LSSVR (seasonal autoregressive integrated moving average-particle swarm optimization-least squares support vector regression) model and a SARIMA-MetaFA-LSSVR (seasonal autogressive integrated moving average-metaheuristic firefly algorithm-least squares support vector regression) | **Single ML models (ANNs, CART, SVM, and LR) + Ensemble models + Hybrid modes (SARIMA-PSO-LSSVR + SARIMA-MetaFA-LSSVR)** | Forecasting energy consumption time series using actual data | The SARIMA-MetaFA-LSSVR and SARIMA-PSO-LSSVR were both implemented as hybrid models A comprehensive comparison demonstrated that the hybrid model is more accurate than single and ensemble models | RapidMiner Studio, IBM SPSS Modeler, WEKA, and MATLAB |
| 24 | Full-scale smart decision support system (SDSS) | **The DM techniques and TS analysis for identifying energy use patterns at appliance level + dynamic multi-objective optimization algorithm to optimize appliance operating schedule** | 1. Framework for energy-saving decision process is presented2. The framework can serve as a basis for the future development of a full-scale smart decision support system (SDSS) | Specifically, it integrates data analytics and dynamic multi-objective optimization modules for generating energy consumption patterns and alternative energy-saving solutions at home appliance level | MATLAB, the web server (e.g., XAMPP controller), and MySQL |
| 25 | SARIMA-MetaFA-LSSVR (seasonal autogressive integrated moving average-metaheuristic firefly algorithm-least squares support vector regression) | **SARIMA-MetaFA-LSSVR system compared with SARIMA (baseline linear model) + LSSVR (baseline nonlinear model) + MetaFA-LSSVR (optimized nonlinear model)** | Predicting real-time building energy consumption data collected by a smart grid | The proposed system exhibited improved performance measures in the range of 368–1132% compared with the other models | MATLAB |
| 26 | Ensemble models that integrate multiple classifiers, the voting, bagging, tiering, and stacking combination methods | **Individual ML techniques (ANNs, CART, CHAID, LR, GLR, and SVMs) + Ensemble models** | Early prediction of building cooling load (CL) and heating load (HL) | This work also confirms that the ensemble model (SVR + ANN) and SVR substantially improve performance in predicting CL and HL | IBM SPSS Modeler |
| **Office Space** | 27 | Hybrid ARIMA—MetaFA–LSSVR | **Standard (ARIMA + LSSVR) +** **ARIMA—MetaFA–LSSVR** | Forecast Electricity Consumption of Smart Grid-Based Air Conditioners | The proposed ARIMA–MetaFA–LSSVR model was superior to the others in terms of performance measures (i.e., RMSE, MAE, and R) in predicting 1-day-ahead electricity consumption of air conditioners | MATLAB |
| 28 | Neural network auto regressive (NNAR) | **Standard ARIMA + hybrid ANN–ARIMA + two-sigma rule** | Detecting anomalous patterns in large data sets for real-time of building office space energy consumption | The comparison results confirm that the hybrid NNAR method obtains more accurate predictions of electricity consumption compared to standard ARIMA | R Language |
| **HVAC Facilities** | 29 | Evolutionary multivariate adaptive regression splines (EMARS), a hybrid of the multivariate adaptive regression splines (MARS) and the artificial bee colony (ABC) | **BPNN + CART + GP + SVM + EMARS** | Predict the coefficient of performance (COP) for refrigeration equipment under various R404A refrigerant conditions | The EMARS testing values obtained for RMSE, MAE, and MAPE for the vapor phase and liquid phase were, respectively, at least 235 % and 150 % below those of BPNN, CART, GP, and SVM | MATLAB |
| 30 | Generalized linear regression (GLR) | **ANNs + SVMs + CART + MR + GLR + CHAID** |  Predict the coefficient of performance (COP) for refrigeration equipment under various R404A refrigerant conditions | In the liquid leakage phase, ANNs provide the best performance In the vapor leakage phase, the best model is the GLR model | IBM SPSS modeler |
| 31 |  Refrigeration system performance | **ANNs + SVMs + CART + MR + GLR + CHAID** | Predict the coefficient of performance (COP) for refrigeration equipment under various R404A refrigerant conditions | ANNs predicts C.O.P the best, while in the vapor leakage phase, GLR obtains the most accurate value. | IBM SPSS modeler |
| ***Project Estimation and Financial Forecasts*** | **Cost Estimates** | 32 | Genetic-algorithm-based artificial neural network (GA-ANN) | **GLR + GA-NLR + GA-ANN + GA-CBR + ANNs** | Forecast bid award amounts for bridge construction projects | The results showed that the GA-ANN (No 10), with three neurons in the hidden layer and α = 09, achieved the highest performance | Evolver & SPSS Statistics |
| 33 | Hierarchical classification and regression (HCR) | **Single ML models (LR + ANNs + SVR) + Ensemble model (HCR)** |  Predict the cost of TFT-LCD inspection and repair equipment | The experimental results show that the proposed HCR approach to developing prediction models outperforms single flat regression models in terms of MAPE and RMSE indicators | WEKA |
| 34 | Evolutionary support vector machine inference model (ESIM) | **ESIM hybrid model (Integrating SVM with a fast messy genetic algorithm (fmGA))** | Predict the cost of manufacturing thin-film transistor liquid–crystal display (TFT-LCD) equipment | The investigation demonstrates that the proposed ESIM can accurately predict the costs of TFT-LCD fabrication equipment (Marco, LOT, and Repair systems) | MATLAB |
| 35 | Incorporates MRA with GRNN (SMRA–GRNN), and GRNN with CBR (GRNN–CBR) | **MRA + ANNs + CBR + Hi (MRA-GRNN + GRNN-CBR + CBREWA-GRNN-CBR + GRNN-CBRA)** | Forecast costs of thin-film transistor liquid-crystal display (TFT-LCD) equipment | The proposed hybrid AI-based technique was up to 354% better than that to the conventional approach currently used in this sector | Neural Tool, Excel macros, and SPSS Statistics |
| 36 | AHP-based reasoning (CBR) | **EW-CBR + EM-CBR** | Estimate Pavement Maintenance Cost | The CBR model was found to be more effective in considering the experience-based weights of these attributes compared to the situation when it treated them as equally important | SPSS + Web programming |
| 37 | Multivariate regression | **Multivariate regression + goodness of ﬁt + statistical validity** | Cost Estimates for Highway Project | The results of the statistical analysis show a strong relationship between the item quantities and the parameters adopted in the models |  SPSS Statistics. |
| **Effort Estimates** | 38 | Multiple regression and backward propagation ANNs | Multiple **regression + ANNs** | Estimate the person-hour of an ERP system development project efficiently | Logarithmic stepwise regression model is superior to ordinary multiple regression and backward propagation ANNs in addition to the accuracy level and explicit explanation |  PASW modeler |
| 39 | Evolutionary support vector machine inference model (ESIM) | **Single ML (SVR + ANNs) + Hybrid model (Integrating SVM with a fast messy genetic algorithm (fmGA)** | Estimating the person-hour of ERP system development projects | Through the cross-validation and prediction power testing, ESIM and ANNs show better performance than conventional SVR in ERP project effort prediction | MATLAB  |
| **Time Estimates** | 40 | (略) |  |  |  |  |
| 41 | (略) |  |  |  |  |
| ***Optimization/Classification Modeling in Project Management and Engineering Applications*** | **Optimization Modeling** | 42 | Adaptive multiple objective symbiotic organisms search (AMOSOS) | **AMOSOS** | Make Time-Cost Tradeoffs in Repetitive Project Scheduling Problem | The Pareto front that was generated by AMOSOS provides information that helps decision makers in construction projects optimally trade-off the two important considerations of duration and cost | MATLAB |
| 43 | Dijkstra's algorithm | **Dijkstra's algorithm is used as the path-planning algorithm** | Optimal path planning in real time for dynamic building fire rescue operations using wireless sensors and visual guidance | 1. The study revealed that the rescue path-system can operate optimally and efficiently when smoke sensors are properly installed2. Bluetooth sensors detected temperature and information about [smoke](https://www.sciencedirect.com/topics/engineering/smoke) around them. | C# (pronounced C sharp), a programming language in Microsoft Visual Studio, SQL server, MySQL databases, [building information modeling](https://www.sciencedirect.com/topics/engineering/building-information-modeling) (BIM) [[25](https://www.sciencedirect.com/science/article/pii/S0926580518307143?via%3Dihub" \l "bb0125)]. The PHP webpages served and mobile phone apps |
| 44 | Modified firefly algorithm (MFA) | **Logistic and Gauss/mouse chaotic maps, adaptive inertia weight, and Lévy flight with a conventional firefly algorithm (FA)** | Multidimensional structural design optimization | The comparison confirmed that the proposed MFA outperformed the conventional FA and the other algorithms in terms of average optimal values and the standard deviations of most considered benchmark functions | MATLAB |
| 45 | Genetic algorithm in SEM procedure (GA-SEM) | **GA + SEM** | Explore the causal relationship between TTS usage and construction engineering project performance (PP) | The analytical results of the demonstrated model show that not all the project management body of knowledge (PMBOK) techniques/tools/skills (TTS) have significant stacking effect on construction engineering project performance (PP) but it does have mutual inter-correlations between some PMBOK | MATLAB |
| 46 | Multiobjective optimization model, MUST |  **MUST+ +SPEA2+ LINGO** | Optimization for manpower assignment in consulting engineering firms | The application of MUST in a numerical case verified its performance, with respect to both effectiveness and efficiency The performance has also been highlighted by comparisons with SPEA2 and LINGO |  Programming language |
| **Classification Modeling** | 47 |  Decision tree (C50) | **ANNs + SVM + Bayes Net + C&RT + CHAID + QUEST + C50+ GRI** | Early notice the dispute handling methods in public infrastructure projects | Using the Y1 model and cross-validation, we found that the decision tree C50 model had the highest accuracy of 8392% | IBM SPSS Clementine 120 |
| 48 | fmGA-based SVM (GASVM) | **CART + QUEST + C50 + CHAID + Fast messy genetic algorithm (fmGA) - support vector machine (SVM)** | Provide the proactive-warning and decision-support information needed to manage potential disputes | Compared to the baseline models (i.e., C50, CHAID, CART, and QUEST) and previous work, GASVM provides 589–1295% higher classification accuracy | IBM SPSS modeler, MATLAB |
| 49 | The hybrid models combining multiple MLP classifiers and multi-ple DT classifiers | **MLP - DT**  |  Early prediction of dispute occurrence using conceptual project information as model input. | This study comprehensively compared the effectiveness of various machine learning techniques The combination with multiple MLP classifiers and multiple DT classifiers outperforms other hybrid models, achieving prediction accuracy of 9708% and 9577%, respectively |  |
| 50 | fmGA - SVMs - FL | **SVMs + fmGA + FL** | Predicting project dispute resolution (PDR) outcomes (i.e., mediation, arbitration, litigation, negotiation, and administrative appeals) when the dispute category and phase in which a dispute occurs are known during project execution | The hybrid models exhibited their superiority, stability, efficiency, and ability to avoid overfitting with better accuracy than the single flat SVMs model A 1208% improvement was achieved with one additional AI technique combined with SVMs, while the improvement was 2476% with two AI techniques | MATLAB |
| 51 | Ensemble model (SVMs-ANNs-C50) | **Single model (ANNs +SVMs +DL +TAN +CART +QUEST +C5.0 +CHAID +DA +LR) and ensemble model** | Provide proactive-warning and decision-support information needed for managing potential disputes before disputes occur | In terms of overall performance measurement score (S), SVMs (0781), ensemble model (0773), and ANNs (0751) are the three best classification models | IBM SPSS Modeler |
| 52 | Ensemble models (QUEST + CHAID + C50) | **Single ML techniques (ANNs, CART, CHAID, QUEST, TAN, C50 and SVMs) + Ensemble models** | Predicting dispute resolution outcomes (i.e., mediation, arbitration, litigation, negotiation, administrative appeals or no dispute occurred) | Analytical results exhibit that the combined technique of QUEST + CHAID + C50 has the best classification accuracy at 8465% | IBM SPSS Modeler |
| ***Project Management Information System/Decision Support System*** | **Standalone System** | 53 | [Smart firefly algorithm (SFA)] and machine learning [least squares support vector regression (LSSVR)] + Smart artificial firefly colony algorithm‐based support vector regression (SAFCA‐SVR) | **SAFCA‐SVR + SFA LS-SVR** | Solve numerous civil engineering problems (energy-efficient buildings, construction material strength, concrete structure shear strength, bridge scour depth, and subbase soil modulus) | The experiments in this study confirmed that the proposed nature-inspired metaheuristic regression system can assist civil engineers and construction managers in assessing, benchmarking, diagnosing, tracking, forecasting, and simulating engineering data for knowledge generation and decision making | MATLAB |
| 54 | [Smart firefly algorithm (SFA)] and machine learning [least squares support vector machine (LSSVM)] | **SFA-LSSVM** | Provide decision-makers with timely warnings of geotechnical hazards | The SFA was integrated with the LSSVM to create a metaheuristic optimized classification modelA GUI was developed for the SFA–LSSVM model to enhance its usability for new users | MATLAB |
| 55 | Natural logarithm of the quantity model with the original measurement unit (REQ) |  **Linear and log-linear statistical approaches were adopted to create most advantageous models,** | Facilitate information management and generate preliminary budgets for transportation agencies | This paper presents a practical data-mining process starting from collection of data, preprocessing, and construction of a novel project cost data warehouse A parametric prediction technique is applied to establish useful estimation models | Microsoft Access, Visual Basic, and structured query language (SQL) |
| 56 | Generalized linear model (GLM) | **GLM** | Estimating the cost of transportation projects | To transform the estimating models into user-friendly applications, a prototype Preliminary Item-level Cost Estimating System (PILCES) was developed with the aid of computer programming |  |
| **Web-based System** | 57 | SARIMA-MetaFA-LSSVR (seasonal autogressive integrated moving average-metaheuristic firefly algorithm-least squares support vector regression) | **SARIMA-MetaFA-LSSVR** | Identify anomalous energy consumption in a home in real time | The web-based system was designed to provide an attractive and user-friendly interface for advanced data mining | MATLAB, Spring Framework, IntelliJ Idea software and Bootstrap |
| 58 | Correlation method and K-means algorithm | **correlation method + K-means algorithm** | Provides a real-time visualization of anomalous consumption based on data from smart meters and sensors to various stakeholders | The web-based system can create graphs of Electricity Consumption so that users can visualize their energy usage status in real time | PHP, Apache Web server, and the relational database MySQL |
| 59 | Visualized EVM system | **EV + DBMS + MCDM** | To monitor project progress and assess project achievements by converting project data into manageable information clusters | The EVMS is a Web-based system via which owners, executive managers, project managers and engineers can access real-time construction project data and progress reports with distinct authorized accounts | JavaScript, PHP (PHP Hypertext Preprocessor), and Open Flash Chart for the client-side, server-side language, and flash graph drawing tool |
| 60 |  Web-based CBR system | **EW-CBR + EM-CBR** | Determine preliminary project cost with readily available information rapidly based on previous experience of pavement maintenance related construction to assist decision makers in project screening and budget allocation | The world wide web makes the CBR an powerful tool with the following characteristics | Relational database management system (RDBMS), MySQL, the web server used Apache, and PHP 50 |
| 61 | Web-based preliminary item-level cost estimating system (WBPILCES) | **RDBMS + Crow's Foot** | Develop a system architecture for a preliminary cost estimation system that toggles project input information, predictive item-level quantity, and segregates unit price of highway projects | The system has proven to narrow down the range of preliminary cost estimates and made these cost estimates more accurate | PHP (PHP Hypertext Preprocessor), APACHE server, and MySQL database server |