

Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v39n6p729-736/2019>

K-NEAREST NEIGHBORS METHOD FOR PREDICTION OF FUEL CONSUMPTION IN TRACTOR-CHISEL PLOW SYSTEMS

Naji Mordi Naji Al-Dosary^{1*}, Saad Abdulrahman Al-Hamed¹, Abdulwahed Mohamed Aboukarima^{1,2}

^{1*}Corresponding author. King Saud University-College of Food and Agriculture Sciences/ Riyadh, Saudi Arabia.
E-mail: nalsawiyah@ksu.edu.sa | ORCID ID: <https://orcid.org/0000-0001-9459-4206>

KEYWORDS

Machine-learning algorithms, tillage, prediction, k-nearest neighbors, fuel consumption, chisel plow.

ABSTRACT

Most important farm operations require a significant amount of energy, and this consumes a major portion of the farm's budget. Consequently, analyzing the fuel consumption of agricultural machinery for farm operations of different sizes makes it possible to predict fuel consumption to set an appropriate budget for energy. The main purpose of this study was to determine the ability of the k-nearest neighbors (KNN) algorithm to predict the fuel consumption of tractor-chisel plow systems correctly. A training-set design of 139 points of 173 data points obtained from the literature was utilized, and the remaining 34 data points were applied as a test set. The input parameters were tractor power, plowing width, depth and speed of plowing, soil percentages of sand, silt, and clay, initial soil moisture content, and initial soil bulk density. The predictive power of the KNN method was compared with that of multiple linear regression (MLR), and experimental data were used to determine the predictive power of both methods. The KNN method generated better results than the multiple linear regression method. The test dataset correlation coefficients were 0.817 for the KNN ($k = 2$) method and 0.422 for the multiple linear regression method. This study suggests that the KNN method with $k = 2$ (two nearest neighbors) is suitable for estimating the fuel consumption of tractor-chisel plow systems for input values within the studied range.

INTRODUCTION

Developing the ability to predict fuel consumption of tractor-machinery systems is extremely beneficial for farms for budgeting and management; however, fuel consumption is measured by the amount of fuel used during a specific time period (Grisso et al., 2010). Furthermore, efficient planning of mechanized farming operations is a complex task, because it involves multiple factors related to the soil composition, the implemented machine, and the decision-making personnel (Borges et al., 2017). Additionally, predicting tractor fuel consumption can lead to more decisions that are appropriate for tractor management (Karparvarfard & Rahmanian-Koushkaki, 2015). Thus, predictive models capable of forecasting the fuel consumption of tractor-machinery systems under different conditions can help farmers optimize their fuel expenditure. Considerable research has addressed the prediction of fuel consumption during the tillage of selected regions with solutions using traditional, statistical, and modern computational methods, or combinations among those literature results (Karparvarfard & Rahmanian-Koushkaki,

2015; Tayel et al., 2015; Almaliki et al., 2016; Borges et al., 2017; Ranjbarian et al., 2017). Successful prediction of the fuel consumption of tractor-chisel systems can aid in selecting tractors that minimize the cost of fuel for tillage.

The chisel plow is considered a primary tool for tillage, because it is mainly used in initial soil working operations (Kheiry et al., 2017). The performance parameters of chisel plows include measurements of draft, drawbar power, actual field capacity, field efficiency, and fuel consumption rate (Bashir et al., 2015). However, prediction of fuel consumption is one of the most challenging tasks; thus, scientists, manufacturers, and users have shown great interest in developing methods that would predict fuel consumption. Although many algorithms have been proposed, accurate prediction of fuel consumption during tillage continues very difficult. Because the price of diesel fuel is high, the ability to predict fuel consumption accurately is potentially advantageous for controlling the cost of crop production and would enable farmers to adjust their equipment for optimal fuel utilization.

¹ King Saud University-College of Food and Agriculture Sciences/ Riyadh, Saudi Arabia.

² Agricultural Engineering Research Institute-Agricultural Research Center/ Giza, Egypt.

Received in: 2-27-2019

Accepted in: 9-9-2019



Tillage is among the most important farm operations, because it requires significant energy and consumes a major portion of the farm's energy budget (Rashidi et al., 2013; Mari et al., 2014). Farmers typically record large numbers of data related to the operation of agricultural machinery, and processing, analyzing, and retrieving significant information from this abundance of farm machinery data are necessary. Utilization of information and modern computational methods, such as machine-learning algorithms, provides knowledge and reveals trends related to the rate of fuel consumption, which in turn affects the choice of appropriate tractor–implement systems and operating conditions, thus reducing the cost of production.

Machine-learning methods are used to solve problems in which the relationship between input and output variables is not known or is difficult to derive. The “learning” term denotes the automatic acquisition of structural descriptions from examples of what is being described (McQueen et al., 1995). Unlike traditional statistical methods, machine-learning methods do not make assumptions about the correct structure of the data model that describes the data. This characteristic is useful for modeling complex nonlinear behavior (Gonzalez-Sanchez et al., 2014). Many methods, such as artificial neural networks (ANNs), radial basis function networks (RBFs), k-nearest neighbors (KNN), and self-organizing maps (SOMs) have been developed for prediction of time series given large datasets with many explanatory variables. The KNN method can be used on nonlinear data for which classical assumptions cannot be made. The KNN method is considered a simple method for analysis of multidimensional data (Alkhatib et al., 2013). Although this method is simple, it is advantageous compared with other methods, allowing the user to generalize based on relatively small training sets (Rokach, 2010).

Originally, KNN was used for classification; however, in the past few decades, this method has also been used for prediction. In the classification approach, a dataset is divided into training and testing datasets. The KNN method uses a similarity measure for comparing the testing data with training data. For prediction of output variables, it chooses k data points from the training dataset that are close to the testing dataset. It is also regarded as a lazy learning method, which does not build a model or a function, but yields the closest k records of the training dataset that are the most similar to the points that are to be categorized (Alkhatib et al., 2013). In the KNN approach, it is especially important to choose the number of KNNs properly, because this choice can strongly affect the predictive power of the method. Small values of k lead to overfitting (high variance), while large values of k result in very biased models. For example, the KNN method has been used for weather prediction. The system generated is relatively accurate for forecasts for months into the future (Jan et al., 2008). The performances of the nearest neighbors (IBk), regression by discretization, and isotonic regression classifiers were compared for the prediction of predefined precipitation classes over Voi, Kenya (Mwagha et al., 2014). The study revealed that the nearest-neighbors method is suitable for prediction of precipitation, given historic rainfall data as a training set. Also, the predictive accuracies with respect to crop yield of multiple linear regression, M5-Prime regression trees, perceptron multilayer neural networks, support vector regression, and

KNN methods were compared (Gonzalez-Sanchez et al., 2014). Real data for an irrigation zone in Mexico were used for building the models. The models were tested on two consecutive-year samples, and the M5-Prime and KNN methods yielded the lowest average root mean square errors (RMSEs). To assist investors, management, decision makers, and users in making correct and informed investment-related decisions, the KNN algorithm and nonlinear regression were also applied to predict stock prices for a sample of six major companies listed on the Jordanian stock exchange. The results showed that the KNN algorithm was robust, with a small error ratio; consequently, the results were rational and also reasonable (Alkhatib et al., 2013).

A KNN classifier was utilized for prediction of daily energy consumption by analyzing historical data on hourly consumption of 520 apartments in Seoul, Republic of Korea. The data were divided into training and testing subsets, with different training and testing ratios, and different qualitative and quantitative measures were used to determine the performance and efficiency of the predictor. The highest accuracy of 95.96% was obtained for the 60%/40% training/testing-set ratio (Wahid & Kim, 2016).

In light of increasing diesel fuel prices, and because existing predictive models of fuel consumption are not satisfactory, collecting large amounts of data on tractor–machinery systems is crucial for economic and farm management analysis. The main objective of this research was to evaluate the machine-learning KNN method for prediction of fuel consumption for different tractor sizes when carrying an implement (chisel plow) in different specifications. The input parameters were tractor power, plowing width, plowing depth, plowing speed, sand, silt, and clay percentages in the soil, initial soil moisture content, and initial soil bulk density. In addition, the KNN predictive ability was compared with that of the multiple linear regression method.

MATERIAL AND METHODS

Construction of the fuel consumption rate model

The fuel consumption rate model was constructed using a machine-learning algorithm — KNN (IBk) — and it was implemented in the WEKA environment. WEKA is an application for performing data-mining tasks that was originally developed at the University of Waikato in New Zealand (Hall et al., 2009). It contains a large collection of state-of-the-art machine-learning and data-mining algorithms written in Java. WEKA has been widely used for many purposes and contains tools for regression, classification, clustering, association rules, visualization, and data preprocessing (Naik & Samant, 2016). The Explorer is the main graphical user interface of WEKA. WEKA has six different panels, accessed by the tabs at the top, which correspond to the various data-mining tasks that are supported by WEKA. In the Preprocess panel, data can be loaded from a file or extracted from a database using an SQL query. The data file can be in the CSV format, or in the system's native ARFF file format. Once a dataset has been read, various data preprocessing tools called “filters” can be applied. The input parameters are the tractor power, the plowing width, the depth and speed of plowing, the soil percentages of sand, silt, and clay, the initial soil moisture content, and the initial soil bulk density. Through the

Explorer's second panel called "Classify", classification and regression algorithms can be applied to the preprocessed data. This panel also enables users to evaluate the resulting models, both numerically, through statistical estimation, and graphically, through visualization of the data and examination of the model (if the model structure is amenable to visualization). Users can also load and save models. In this study, the KNN algorithm was evaluated for its default parameters defined in the WEKA application, except that the k value was varied from 1 to 5.

KNN method

The KNN algorithm is a machine-learning algorithm that is considered a lazy learning algorithm, with a low computational cost and very simple implementation (Alkhatib et al., 2013). It supports classification and regression problems. When making a prediction, it stores the entire training dataset and queries it to locate k data points in the training set that are most similar to the data point to be classified. Therefore, there is no model other than the raw training dataset, and the only computation performed is querying of the training dataset.

When the KNN method is used for regression, the response value is calculated as a weighted sum of the responses of all the k neighbors, where the weight is inversely proportional to the distance from the input record. This distance is generally Euclidean. The Euclidean distance function is defined by Wilson & Martinez (2000) as follows.

$$E(x, p) = \sqrt{\sum_a^m (x_a - p_a)^2} \quad (1)$$

Where,

x and p are the query point and a case from the set of examples, respectively, while m is the number of input variables (attributes).

The algorithm is sensitive to the selection of KNNs. The KNN method has several attractive properties. Beyond the choice of KNNs and the distance metric, no optimization or training is required (Hand et al., 2001). The method takes full advantage of local information and can yield highly nonlinear and highly adaptive decision boundaries. The method's disadvantages are its high computational and memory costs, because all the available data points (i.e., samples) should be scanned to determine the most similar neighbors. The calculation of distances becomes more problematic for higher-dimensional datasets. Despite these issues, the method is popular because of its ease of implementation and the above-mentioned properties (Gonzalez-Sanchez et al., 2014).

When using the KNN algorithm, the dataset should be divided into two subsets: the training dataset, on which the algorithm bases its predictions; and the testing dataset, which is used to test the algorithm's performance on previously unseen data (Imandoust & Bolandraftar, 2013). The training dataset is divided into vectors; then, for each

point in the testing dataset, the distance from the data point to its neighbors in the training dataset is calculated using the Euclidean distance measure in the WEKA tool (Vainionpää & Davidsson, 2014). In the present study, the training dataset contained 139 instances and the testing dataset contained 34 instances.

After selecting the value of k , predictions based on the KNN examples can be made; however, a prediction is the average over the outcomes for KNNs, as specified in [eq. (2)] (Imandoust & Bolandraftar, 2013).

$$y = \frac{1}{k} \sum_{i=1}^k y_i \quad (2)$$

Where,

y_i is the i^{th} example, and

y is the prediction (outcome) for the query point.

Multiple linear regression

Multiple linear regression (MLR) was applied using WEKA. The linear regression captures the variation in the fuel consumption as a function of tractor power, plowing width, depth and speed, soil percentages of sand, silt, and clay, initial soil moisture content, and initial soil bulk density.

Collection of data

Building and testing the KNN fuel consumption rate model

To develop the KNN fuel consumption rate model, from prior literature, any available datasets were mined for tractor–chisel plow systems that are directly related to the subject (El Banna & Helmy, 1992; Abd El Motaleb & Helmy, 1993; Abd El Wahab, 1994; Al-Taieb, 1998; Gomaa, 1998; Abd Alla et al., 1999; El Raie et al., 1999; Metwally et al., 2000; Younis et al., 2000; Badawy et al., 2001; El Sayed & El Kilani, 2002; Al-Jebory, 2011). The collected data were from field experiments in which different chisel plows were used (only one pass over soil) in different sites with different moisture levels, bulk densities, and textures, and with different changeable working conditions. The dataset contained 173 instances, each with nine attributes. The data were randomized and divided it into two datasets. The first dataset, consisting of 139 data points (inputs and output), was used as a training dataset, and the remaining 34 data points were utilized as a testing dataset. The output variable (Y) in the present study was the rate of fuel consumption. The input variables in this study were tractor power, plowing width, plowing depth and plowing speed, soil percentages of sand, silt, and clay, initial soil moisture content, and initial soil bulk density. Descriptive statistics for the collected literature data on the fuel consumption rates of tractor–chisel plow systems are listed in Table 1. Meanwhile, descriptive statistics for performance parameters that were used as inputs to predictive models are listed in Table 2.

TABLE 1. Descriptive statistics for the fuel consumption rate data of a tractor–chisel–plow system, collected from the literature.

Statistic	Value	Unit
Minimum	2.4	Lh ⁻¹
Maximum	20.8	Lh ⁻¹
Average	13.9	Lh ⁻¹
Median	14.4	Lh ⁻¹
Kurtosis coefficient	-0.2	Dimensionless
Skewness coefficient	-0.7	Dimensionless
Coefficient of variation (CV)	31.3	%

Accuracy metrics

To evaluate the accuracy of the predictive model, different accuracy metrics were used: the correlation coefficient (R), mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE). These metrics collectively constituted the WEKA result panel. The four metrics were defined as follows.

For the metrics definitions, if Y_t is the actual observation for period t and F_t is the prediction for the same period, the correlation coefficient determines the linear relationship between the two variables. It is defined in [eq. (3)] (Makridakis et al., 1998):

$$R = \frac{\sum_{t=1}^n (Y_t - \bar{Y}) \times (F_t - \bar{F})}{\sqrt{\sum_{t=1}^n (Y_t - \bar{Y})^2} \times \sqrt{\sum_{t=1}^n (F_t - \bar{F})^2}} \quad (3)$$

The correlation coefficient takes values from -1 to +1. A positive correlation coefficient implies that the two variables vary in the same direction with respect to their means. A negative correlation coefficient implies that the two variables vary in opposite directions with respect to their means. A value close to 0 implies that the two variables have little linear dependency. The error e_t between the two variables is defined as:

$$e_t = Y_t - F_t \quad (4)$$

If there are observations and predictions for n periods, then there are n error terms, and the following standard statistical measures can be defined as shown in eqs 5-8:

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (6)$$

$$RRSE = \sqrt{\frac{\sum_{t=1}^n (Y_t - F_t)^2}{\sum_{t=1}^n (F_t - \bar{F})^2}} \times 100 \quad (7)$$

$$RAE = \frac{\sum_{t=1}^n |Y_t - F_t|}{\sum_{t=1}^n |F_t - \bar{F}|} \times 100 \quad (8)$$

Here, MAE is the mean absolute error, and it refers to the sum of individual absolute errors normalized by the number of samples. The quantity RMSE is the root mean squared error, and it is a modification of the mean absolute value, with the absolute value of an individual error term replaced with a square. Both MAE and RMSE measure the average difference between the predicted and actual values. However, RMSE is more commonly used to measure the model’s goodness of fit. RMSE pays more attention to large errors, owing to its square term. The RAE measure calculates the variance of the model when units are not important when comparing models. The RRSE measure compares the model prediction against the mean. For this metric, a value below 100% indicates a better performance than the average.

TABLE 2. Descriptive statistics for the performance parameters used as inputs to predictive models.

Performance parameters	Min.	Max.	Ave.	Med.	Kurtosis (dimensionless)	Skewness (dimensionless)	CV (%)
TP (kW)	25.35	104.40	68.26	59.66	-1.19	0.27	34.92
PW (m)	1.05	3.10	1.96	1.75	1.21	1.60	23.44
PD (cm)	10.00	30.00	16.87	15.00	-0.35	0.62	33.21
PS (kmh ⁻¹)	1.06	6.92	3.90	3.60	0.33	0.65	32.72
Sand (%)	11.38	55.71	32.91	30.92	-0.21	0.60	38.48
Silt (%)	15.60	55.20	26.82	24.43	0.66	0.91	35.73
Clay (%)	19.07	53.74	40.27	41.99	-0.50	-0.47	19.98
MC (% db)	10.42	50.20	20.97	21.19	14.00	3.33	29.49
BD (gcm ⁻³)	1.15	1.86	1.35	1.34	3.15	1.51	10.43

TP: tractor power, PW: plowing width, PD: plowing depth, PS: plowing speed, MC: initial soil moisture content, BD: initial soil bulk density, Min.: minimum, Max.: maximum, Ave.: average, Med.: median.

RESULTS AND DISCUSSION

Data analysis

The general statistical characterization of fuel consumption rate data, which illustrates the descriptive statistics of fuel consumption data for a dataset of 173 data points, is shown in Table 1. The small difference between the average and the median values of fuel consumption rates indicates that the distribution of data is close to normal. Skewness is a lack of symmetry in a probability distribution, and kurtosis is the measurement of separation of smoothing probability distribution from a normal distribution shape (Everitt & Skrondal, 2010). In Table 1, for fuel consumption data, the skewness coefficient is a negative value, which indicates that data are skewed left; also, the kurtosis coefficient is a negative value, which indicates that data have a platykurtic distribution. However, Borges et al. (2017) also obtained negative values for skewness and kurtosis for tractor fuel consumption data. In addition, the variation coefficient is slightly high (31.3%), because the aforementioned fuel rate data were collected from different sources. The maximal fuel rate was 20.8 Lh^{-1} , and the minimal was 2.4 Lh^{-1} ; such a wide scatter indicates that different parameters likely affect the rate of fuel consumption.

Table 2 lists the descriptive statistics of the performance parameters used as inputs to the predictive models corresponding to the 173 data points in the dataset. In this table, slight differences between the averages and medians of parameters can be seen for the plowing speed, initial soil moisture, and initial soil bulk density content parameters. Additionally, the kurtosis and skewness coefficients are different values between negative and positive. However, the values for asymmetry and kurtosis between -2 and $+2$ are considered acceptable to prove a normal univariate distribution (George & Mallery, 2010). The Excel software was used to calculate skewness and kurtosis. Table 2 also shows that the tractor power, plowing depth, plowing speed, silt percentage, sand percentage, and initial soil moisture content variables have the highest variation coefficients, because these data are collected from different sources, under different experimental conditions.

Performances of the KNN and multiple linear regression algorithms

The simplest KNN method assumes $k = 1$. With this value, the predictive power of the model is rather unsatisfactory; because the model is characterized by high variance, it overfits the training-set data and performs poorly on the testing-set data. Increasing the value of k reduces the variance but may increase the bias. Thus, the algorithm is sensitive to the selection of k (Hand et al., 2001). In this study, the number of KNNs was varied from 1 to 5. For the testing set, the results showed that the correlation coefficient of prediction was high for $k = 2$, as shown in Figure 1. The KNN method generated better results than the multiple linear regression method. The test dataset correlation coefficients were 0.817 for the KNN method and 0.422 for the multiple linear regression method with $k = 2$ (two nearest neighbors). Meanwhile, the MAE, RMSE, RAE, and RRSE were small for $k = 2$, as shown in Figures 2, 3, 4, and 5, respectively.

Table 3 tabulates the results obtained for the different algorithms on the training-set data. The purpose was to compare comprehensively the performances of multiple linear regression with those of various KNN algorithms with different values of k . Clearly, the KNN algorithm yields a significantly higher prediction accuracy of fuel consumption, compared with multiple linear regression. Comparing the correlation coefficients reveals the ability of the KNN algorithm to improve the accuracy of fuel consumption prediction, as indicated by noticeable reductions in the MAE, RMSE, RAE, and RRSE measures, as shown in Table 3. This implies that predictions generated by the KNN algorithm-based model exhibit a relatively small deviation from the actual fuel consumption data.

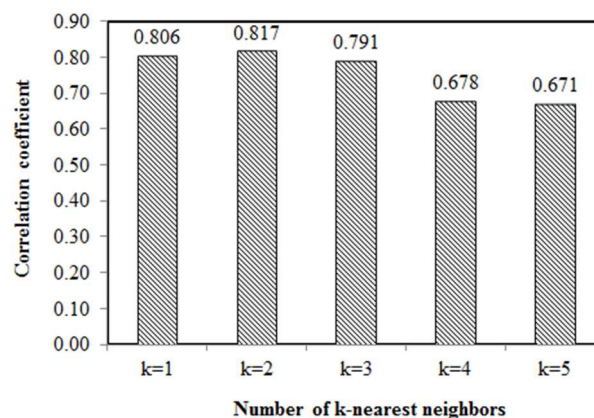


FIGURE 1. Effect of the number of k-nearest neighbors on the correlation coefficient.

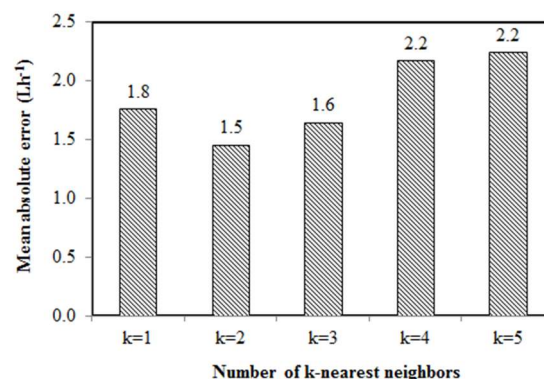


FIGURE 2. Effect of the number of k-nearest neighbors on the MAE.

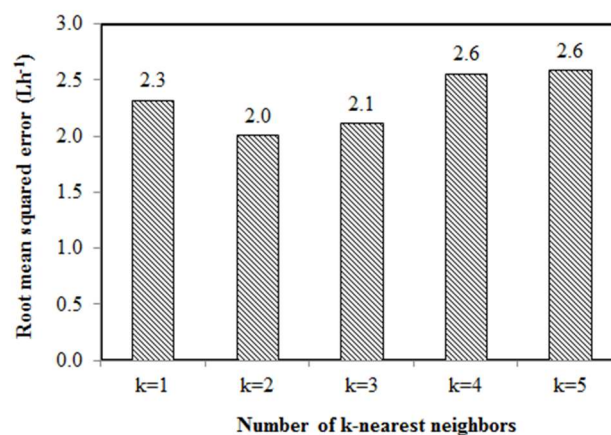


FIGURE 3. Effect of the number of k-nearest neighbors on the RMSE.

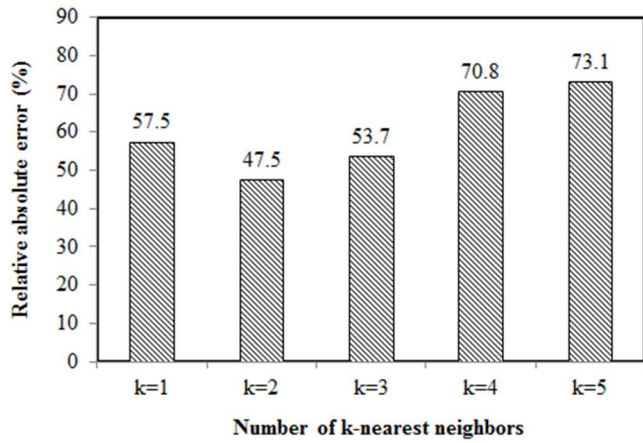


FIGURE 4. Effect of the number of k-nearest neighbors on the RAE.

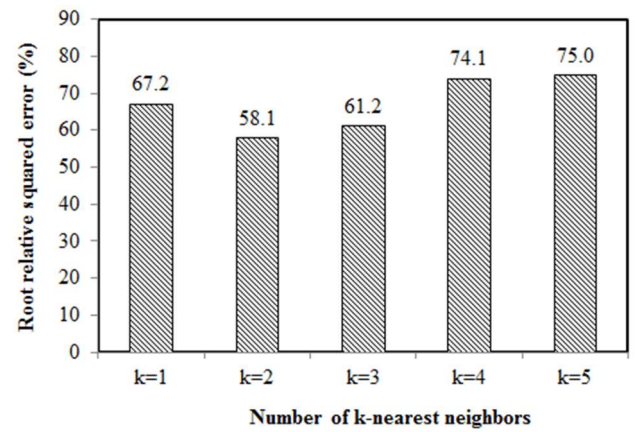


FIGURE 5. Effect of the number of k-nearest neighbors on the RRSE.

TABLE 3. Errors of different k-nearest neighbors on the training dataset.

Number of k-nearest neighbors	Correlation coefficient	MAE (Lh ⁻¹)	RMSE (Lh ⁻¹)	RAE (%)	RRSE (%)
1	1	0	0	0	0
2	0.934	1.07	1.63	28.14	35.84
3	0.909	1.28	1.90	33.86	41.82
4	0.883	1.54	2.14	40.62	47.10
5	0.848	1.78	2.42	46.89	53.28
Multiple linear regression	0.576	3.12	3.71	82.18	81.74

Figure 6 shows the relationship between the actual and predicted fuel consumption rates for data in the testing set, while Table 4 lists the straight-line equations of the "slope–intercept" form ($Y = mX + b$) for estimating the fuel

consumption rate yielded by the KNN method for different k (X corresponds to the actual fuel consumption rate). It is clear from Figure 6 and Table 4 that the best fit is obtained for $k = 2$, manifested as a near-unity slope of the straight line.

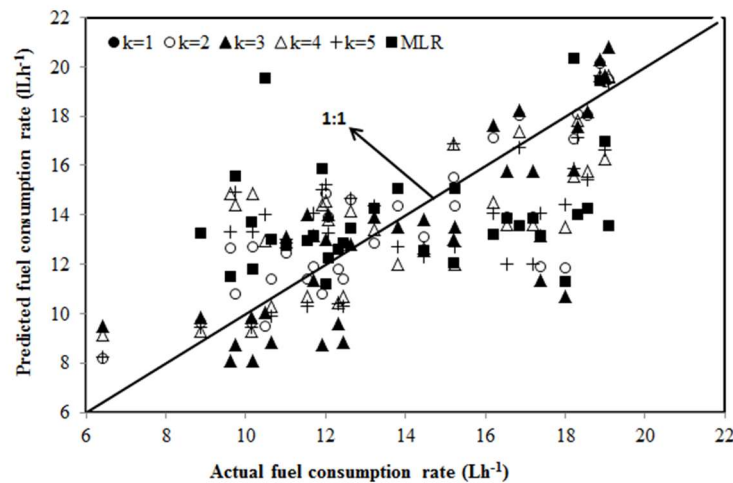


FIGURE 6. Relationship between the actual and predicted fuel consumption rates for data in the testing set, using the KNN and multiple linear regression algorithms (MLR).

TABLE 4. Straight-line equations of the slope–intercept for estimating the fuel consumption rate yielded by the KNN method for different k, for data in the testing set.

Y (fuel consumption, Lh ⁻¹ estimated when k = n)	Intercept (b)	Slope (m)	Correlation coefficient
1	3.870	0.755	0.806
2	1.299	0.911	0.817
3	3.870	0.755	0.806
4	1.718	0.880	0.678
5	2.060	0.859	0.671
Multiple linear regression	6.126	0.559	0.422

CONCLUSIONS

The results of this study confirmed that the successful prediction of fuel consumption by tractor–chisel systems can aid farmers in selecting appropriate tractors and implements, thus helping to reduce the cost of tillage. An efficient k-nearest neighbors (KNN) algorithm with $k = 2$ was adopted to perform such tests on the training dataset. The KNN algorithm was stable and robust, and it exhibited a relatively small error ratio, thus providing rational and reasonable results. Regarding actual fuel consumption, the model predictions were close to actual values. This implies that this data-mining technique can help farmers to select proper equipment and operation parameters to reduce fuel consumption during tillage with chisel plows.

ACKNOWLEDGMENTS

With respect and gratitude, the authors thank the Deanship of Scientific Research, Researchers Support Services Unit, and Agricultural Research Center at the College of Food and Agriculture Sciences, King Saud University for their moral support and encouragement.

REFERENCES

- Abd Alla HE, El Sayed GH, Badr SE (1999) Selecting the proper system for seedbed preparation and sowing method to obtain the highest wheat yield. *Misr Journal of Agricultural Engineering* 16: 663-674.
- Abd El Motaleb IA, Helmy MA (1993) A field performance of front wheel auxiliary drive tractors. *Misr Journal of Agricultural Engineering* 10:805-815.
- Abd El Wahab MK (1994) Minimum tillage by a simple combination. *Misr Journal of Agricultural Engineering* 11:711-724.
- Al-Jebory MA (2011) Effect of two plows, soil moisture and practical speed on some performance parameters and soil physical properties. *AL-TAQANI* 24:A1-A18 (In Arabic).
- Alkhatib K, Najadat H, Hmeidi I, Shatnawi MKA (2013) Stock price prediction using k-nearest neighbor (kNN) algorithm. *International Journal of Business, Humanities and Technology* 3(3):32-44.
- Almaliki S, Alimardani R, Omid M (2016) Fuel consumption models of MF285 tractor under various field conditions. *Agricultural Engineering International: CIGR Journal* 18(3):147-158.
- Al-Taieb AE (1998) Effect of different tillage methods on some physical properties of soil and sunflower yield. *Misr Journal of Agricultural Engineering* 15(1):159-173.
- Badawy ME, El Khateeb HA, Meleha MI (2001) Effect of different seedbed preparation on water requirement and sunflower yield. *Misr Journal of Agricultural Engineering* 18:445-460.
- Bashir MA, Dawelbeit MI, Eltom MO, Tanakamaru H (2015) Performance of different tillage implements and their effects on sorghum and maize grown in Gezira Vertisols, Sudan. *International Journal of Scientific & Technology Research* 4(4):237-242.
- Borges PHM, Mendoza ZMSH, Maia JCS, Bianchini A, Fernandes HC (2017) Estimation of fuel consumption in agricultural mechanized operations using artificial neural networks. *Journal of the Brazilian Association of Agricultural Engineering* 37(1):136-147. DOI: <http://dx.doi.org/10.1590/1809-4430-eng.agric.v37n1p136-147/2017>
- El Banna EB, Helmy MA (1992) Influence of precision tillage system on soil compaction, power requirements and wheat crop yield. *Misr Journal of Agricultural Engineering* 9:537-558.
- El Raie AES, Taiab AZ, Tadros MR, Abeed MMA (1999) Energy requirements for the cultivation of vineyards. *Misr Journal of Agricultural Engineering* 16:101-126.
- El Sayed GH, El Kilani RMM (2002) Predicted vs. measured tractor fuel - consumption in soil-working operations. *Misr Journal of Agricultural Engineering* 19:268-284.
- Everitt BS, Skrondal A (2010) *The Cambridge dictionary of statistics*. New York, Cambridge University Press, 4 ed. 480p.
- George D, Mallery P (2010) *SPSS for Windows step by step: a simple guide and reference 17.0 update*. Pearson, 10 ed.
- Gomaa SM (1998) Effect of blade arrangement of chisel plough on ploughing operation and soil properties. *Misr Journal of Agricultural Engineering* 15:145-158.
- Gonzalez-Sanchez, A, Frausto-Solis J, Ojeda-Bustamante W (2014) Predictive ability of machine learning methods for massive crop yield prediction. *Spanish Journal of Agricultural Research* 12(2):313-328. DOI: <https://doi.org/10.5424/sjar/2014122-4439>
- Grisso R, Perumpral JV, Vaughan DH, Roberson GT, Pitman R (2010) Predicting tractor diesel fuel consumption. *Virginia*, p442-073.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) *The WEKA data mining software: an update*. *ACM SIGKDD Explorations Newsletter*, University of Waikato 11(1):10-18.
- Hand D, Mannila H, Smyth P (2001) *Principles of data mining*. MIT Press. Cambridge, The MIT Press, 546p.
- Imandoust SB, Bolandraftar M (2013) Application of k-nearest neighbor (KNN) approach for predicting economic events: theoretical background. *International Journal of Engineering Research and Applications* 3(5):605-610.
- Jan Z, Abrar M, Bashir S, Mirza AM (2008) Seasonal to inter-annual climate prediction using data mining KNN technique. In: Hussain D.M.A., Rajput A.Q.K., Chowdhry B.S., Gee Q. (eds) *Wireless Networks, Information Processing and Systems. IMTIC 2008. Communications in Computer and Information Science*, 20:40-51. DOI: https://doi.org/10.1007/978-3-540-89853-5_7

- Karparvarfard SH, Rahmanian-Koushkaki H (2015) Development of a fuel consumption equation: Test case for a tractor chisel-ploughing in a clay loam soil. *Biosystems Engineering* 130:23-33. DOI: <https://doi.org/10.1016/j.biosystemseng.2014.11.015>
- Kheiry ANO, Mohamed MA, Omer EA, Rahma AE, and Albahi A (2017) Performance evaluation of Giad chisel plow cp007 under different type of soils. *International Journal of Scientific & Engineering Research* 8:1273-1283.
- Makridakis SG, Wheelwright SC, Hyndman RJ (1998) *Forecasting: methods and applications*. New York, Ed. John Wiley & Sons, 3 ed. 656p.
- Mari IA, Ji C, Tagar AA, Chandio FA, Hanif M (2014) Effect of soil forces on the surface of moldboard plow under different working conditions. *Bulgarian Journal of Agricultural Science* 20(2):497-501.
- McQueen RJ, Garner SR, Nevill-Manning CG, Witten IH (1995) Applying machine learning to agricultural data. *Computers and Electronics in Agriculture* 12:275-293. DOI: [https://doi.org/10.1016/0168-1699\(95\)98601-9](https://doi.org/10.1016/0168-1699(95)98601-9)
- Metwally ME, Abou Shieshaa RR, Kholief RM, Kanany RE (2000) Effect of four different tillage systems and nitrogen sources on wheat production under improved salt affected soil. *Misr Journal of Agricultural Engineering* 17(3):539-554.
- Mwagha SM, Muthoni M, Ochieg P (2014) Comparison of nearest neighbor (IBK), regression by discretization and isotonic regression classification algorithms for precipitation classes prediction. *International Journal of Computer Applications* 96:44-48. DOI: <https://doi.org/10.5120/16919-6729>
- Naik A, Samant L (2016) Correlation review of classification algorithm using data mining Tool: WEKA, Rapidminer, Tanagra, Orange, and Knime. *Science Direct, Procedia Computer Science* 85:662-668.
- Ranjbarian S, Askari M, Jannatkah J (2017) Performance of tractor and tillage implements in clay soil. *Journal of the Saudi Society of Agricultural Science* 16:154-162. DOI: <https://doi.org/10.1016/j.jssas.2015.05.003>
- Rashidi M, Najjarzadeh I, Namin ST, Naserzaeim F, Mirzaki SH, Beni MS (2013) Prediction of moldboard plow draft force based on soil moisture content, tillage depth and operation speed. *American-Eurasian Journal of Agricultural & Environmental Sciences* 13(8):1057-1062. DOI: 10.5829/idosi.aejas.2013.13.08.11022.
- Rokach L (2010) *Pattern classification using ensemble methods*. Singapore: World Scientific Publishing Company, 225p.
- Tayel MY, Shaaban SM, Mansour HA (2015) Effect of plowing conditions on the tractor wheel slippage and fuel consumption in sandy soil. *International Journal of ChemTech Research* 8(12):151-159.
- Vainionpää I, Davidsson S (2014) Stock market prediction using the k-nearest neighbors algorithm and a comparison with the moving average formula. Degree Project in Computer Science DD143X. Available: <https://www.diva-portal.org/smash/get/diva2:771141/FULLTEXT01.pdf>. Accessed: Feb, 2018.
- Wahid F, Kim D (2016) A prediction approach for demand analysis of energy consumption using k-nearest neighbor in residential buildings. *International Journal of Smart Home* 10(2):97-108. DOI: <https://doi.org/10.14257/ijsh.2016.10.2.10>
- Wilson DR, Martinez TR (2000) Reduction techniques for instance-based learning algorithms. *Machine Learning* 38:257-286.
- Younis SM, Nasr GM, Al-Tenbi MNS (2000) Technical and economic studies on mechanization of seedbed preparation for cotton production. *Misr Journal of Agricultural Engineering* 17:55-78.