Computers and Electronics in Agriculture 154 (2018) 296–303

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Prediction models of starch content in fresh cassava roots for a tapioca starch manufacturer in Thailand

Jirachai Buddhakulsomsiri[a], Parthana Parthanadee[b,*], Warut Pannakkong[a]

[a] School of Manufacturing Systems and Mechanical Engineering, Sirindhorn International Institute of Technology, Thammasat University, Thailand
[b] Department of Agro-Industrial Technology, Faculty of Agro-Industry, Kasetsart University, Thailand

ARTICLE INFO

Keywords:
Prediction
Starch content
Regression
Artificial neural network
Deep belief network

ABSTRACT

This paper involves an application of prediction models to study quality of incoming raw materials of a tapioca starch manufacturer in Thailand. The objectives are to estimate starch content of fresh cassava roots and to identify significant factors that affect starch content in cassava roots. Three prediction models, including multiple regression, artificial neural network (ANN), and hybrid deep belief network (HDBN), are implemented. Input data were collected from 242 farmers from 49 different sub-districts in Nakhon Ratchasima province in the Northeast of Thailand, who supply fresh cassava roots to the manufacturing plant. Potential factors are classified into four categories: farmers' demographics, cultivation activities, harvesting activities, and logistics activities, a total of 38 variables. Regression models, ANNs with one hidden layer, and HDBNs were constructed for starch content prediction. Prediction performances were evaluated using the root mean square error (RMSE) and mean absolute percentage errors (MAPE), which were 2.44 percent of starch content and 7.283% for the best regression model; 2.41 and 7.055% for the best ANN, and 2.35 and 6.226% for the best HDBN, respectively. The results indicate that HDBN outperforms the other two models in terms of prediction performance. The final regression model and the best ANN are primarily used to identify seven important factors that can potentially describe starch content. These include harvest age, planting density, growing season, farm location, type of soil, cassava variety, and weed control method.

1. Introduction

Cassava is the third most important economic crop of Thailand, after para rubber and rice (Thailand Ministry of Commerce, 2018). Thailand is the world's second largest cassava root producer, and is the world's largest exporter of cassava products due to its low domestic consumption. Other major cassava producer countries such as Nigeria, Indonesia, and Brazil, produce the roots for domestic consumption. Thailand accounts for 69% of the global market share, followed by Vietnam, while China is the world's leading importer of cassava products (The Office of Agricultural Economics [OAE], 2018a; FAOSTAT, 2018). Fresh cassava roots are generally processed into dried chips, pellets, starches (native and modified), and ethanol, which are further utilized as raw materials or ingredients in various industries. These products (67% in a root-equivalent quantity) are exported to several countries around the world, except ethanol, which is almost 100% used domestically (The Thai Tapioca Trade Association, 2016; OAE, 2018a). From the crop years 2013/14 to 2017/18, the average yield of cassava in Thailand is 22.24 tonnes per ha (The Thai Tapioca Trade Association, 2018). The starch content in fresh root is a critical quality characteristic, as it significantly affects the yield of tapioca processing. Many tapioca starch manufacturers, therefore, set the buying prices of cassava roots based on the starch content measured at the factory gates (Buddhakulsomsiri et al., 2015). Due to high variations in farmers' cultivation, harvesting, and transportation practices, farming area, type of soil, geographic location, as well as climate change, there is a high variability in starch content of fresh cassava roots produced in Thailand. For example, based on the data collected in this study, starch content of incoming fresh cassava from 242 farmers ranges from 16.6% to 34%. With such high variability, it is of great interest to tapioca starch manufacturers to be able to estimate or predict starch content of incoming fresh cassava roots because of its impact on raw material procurement and production planning.

Several researchers have studied the effects of cultivation, harvesting, and post-harvesting practices on starch content and properties. Kayode (1983) found that cassava that grew in Southwestern Nigeria should be planted in May and harvested after 12–15 months to achieve high starch content for industrial purposes. Kawano et al. (1987)

* Corresponding author.
E-mail address: parthana.p@ku.ac.th (P. Parthanadee).

https://doi.org/10.1016/j.compag.2018.09.016
Received 30 May 2018; Received in revised form 17 August 2018; Accepted 12 September 2018

reported the effects of genetic and environmental variability (i.e., harvest age, seasons and growing locations, in terms of altitude) on cassava root dry matter content (RDMC). Asadu et al. (1998) found that both climate and altitude greatly affected the soil fertility and thus the cassava yield. Asaoka et al. (1991) revealed the effects of cassava varieties and harvesting ages or seasons on the physico-chemical properties of cassava starch in Colombia. Similar conclusions were drawn by Sriroth et al. (1999) in Thailand and Tan et al. (2017) in China. However, proper varieties and harvesting ages of roots may be different from region to region. Rojanaridpiched (2007) summarized the factors affecting the starch content in cassava roots, including varieties, soil structure and plant nutrients in the soil, harvesting season, plant cutting before harvesting, and post-harvest handling of roots. Many literatures have suggested that harvesting the roots during the wet season would reduce the starch content, as the plant would move the starch to create new sprouts. This has also been stated in the note by Moreno and Gourdji (2015). Furthermore, pre-harvest pruning of cassava plants reduces the starch content at a rate of 7 g/kg (dry matter basis) per day of the pruning-harvest interval, even though it could reduce the susceptibility to post-harvest physiological deterioration, and thus, extend the shelf life of the roots (Van Oirschot, 2000). Pre-harvest pruning also affects the quality of cassava starch; therefore it should be done within one to three weeks before harvesting for acceptable starch quality (Asaoka et al., 1993; Van Oirschot, 2000). Furthermore a review on the postharvest handling and storage of fresh cassava root and products can be found in Uchechukwu-Agua et al. (2015).

Conducting a study to predict and to identify important factors that affect starch content using experiment in a field is subject to resource, plantation area, cost, and time constraints. This often results in a limited number of factors that can be studied. Rather than experimentation, this paper involves an observational study conducted at a factory, which can collect more data in a relatively shorter time. The motivation of using actual observations from cassava fields is that findings from experimental studies, which are conducted under some controlled environment, may or may not be applicable to uncontrolled environment in farmers' fields. The study employed multiple linear regression, artificial neural network (ANN), and a proposed hybrid deep belief network (HDBN) techniques to find the relationships between cassava starch content and cassava productions and other related factors. The models could be used to predict an average starch content of incoming fresh cassava roots. Also, the findings about important variables can provide a guideline for improving starch content in fresh cassava roots, which would help farmers to increase their starch yields and gain more revenue. In addition, starch manufacturers could benefit from a better tapioca processing yield. This would enhance the efficiency of cassava production and processing along the cassava supply chain.

Research studies in the literature that used regression are numerous, since it is well known and widely used, thus, literature review on regression model usage is omitted to save space. For ANN and DBN, a comprehensive review by Loutfi et al. (2015) of recent works on some data preprocessing algorithms and application areas in food quality monitoring indicates five studies that applied ANN and its variants, and one study that used DBN. These recent studies that applied ANN to forecast certain properties or characteristics of food are discussed as follows. Llave et al. (2011) used an ANN model to forecast the cold spot temperature profile in retort sterilization of starch-based foods. Boccorh and Paterson (2002) developed ANNs for forecasting flavor intensity in blackcurrant concentrates from gas chromatographic data of 37 flavor components. They employed the ANN and partial least squares (PLS) models and compared the results. The forecasting result of overall flavor and intensity from the ANN is better than that of PLS. Sablani and Rahman (2003) compared multiple regression and ANN models for the forecasting of thermal conductivity of food, based on temperature, apparent porosity, and moisture content, as input variables. The result showed that the ANN was more accurate than multiple regression. Kerdpiboon et al. (2006) also used multiple regression and an ANN to forecast the shrinkage and rehydration of dried carrots by using the inputs of moisture content and normalized fractal dimension analysis of the cell wall structure, and found that the optimal ANN models could forecast better than the regression models. For DBN, Längkvist et al. (2013) applied the model to perform classification of meat spoilage. It was reported that DBN can detect meat spoilage with better and faster classification than another approach, the support vector machine (SVM). Lertworasirikul (2008) investigated the drying kinetics of semi-finished cassava crackers from a drying process using a tray dryer. The author compared various models, including a diffusion model, Newton model, Page model, Modified Page model, Henderson and Pabis model, MFNN (Multilayer Feedforward Neural Network), and ANFIS (Adaptive-Network-based Fuzzy Inference System), and concluded that the MFNN model performed the best in forecasting the moisture ratio of the product. The MFNN model was enhanced with Nonlinear-Auto-Regressive with Exogenous input (MFNN-NARX) and used to forecast the moisture content and water activity of the product in Lertworasirikul and Tipsuwan (2008). To the best of our knowledge, there is no previous research that has employed these modelling techniques to predict and/or identify factors that significantly affect starch content in fresh cassava roots based on observational data. Moreover, while regression and ANN are widely used techniques, the proposed HDBN, developed in this paper, is shown to outperform the previous techniques in terms of prediction performance.

2. Methods

2.1. Sampling procedures

The study focused on the cassava cultivation areas in Nakhon Ratchasima, a province that produces the largest cassava supply in Thailand, with 215.84 thousand ha of cassava and 4.93 million tonnes of root production, equivalent to 18.10% of the total production (OAE, 2018b) in Thailand. Data were collected from farmers who sold their cassava roots to a large starch manufacturer in the province during January to April (high harvesting season in Thailand). Cassava's important characteristic is the starch content, which was measured from a 5-kg root sample, that was randomly drawn from incoming farmer trucks. Personal interviews with the farmers were conducted while they were queuing and waiting for the laboratory test results of their starch contents. There were a total of 242 farmers from 49 different sub-districts in Nakhon Ratchasima participating in the study. The collected data are divided into five parts, as follows:

Part 1 involves general demographics including years of experience in cassava cultivation, gender, participation in the membership program of the starch manufacturer, frequency of training on cassava farming techniques, and cassava selling channel (directly or through agricultural cooperatives).

Part 2 contains the logistics activity information including type of delivery truck, amount of load (tonnes), distance travelled and travel time from farm to factory, arrival time at the factory, truck ownership, and root arrangement pattern on the truck (horizontal, vertical, or mixed).

Part 3 is about cassava cultivation including plantation area, size of farm (ha), landscape (lowland, upland, hillside, or flatland), soil type, cassava variety, harvesting age, yield (tonnes/ha), and cultivating month.

Part 4 collects harvesting activity information, including the number of days for harvesting the roots after cutting the trunks to prepare cassava stems for the next crop, harvesting practices (harvesting the entire farm at once vs. harvesting partially each time, and harvesting cassava with vs. without cutting off rootstock), harvesting and material handling tools used, occurrence and amount of rotten cassava, cassava breakages, and harvesting time.

Part 5 includes land preparation and planting operations