OPTIMIZATION SEAWEED DRYING EFFICIENCY USING HYBRID SOLAR DRYERS AND SPARSE ROBUST REGRESSION MODELS

NOUR HAMAD SULEIMAN ABU AFOUNA

UNIVERSITI SAINS MALAYSIA

2025

OPTIMIZATION SEAWEED DRYING EFFICIENCY USING HYBRID SOLAR DRYERS AND SPARSE ROBUST REGRESSION MODELS

by

NOUR HAMAD SULEIMAN ABU AFOUNA

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

April 2025

ACKNOWLEDGEMENT

First and foremost, I offer my deepest gratitude to the Almighty God for His grace, strength, and unwavering faithfulness that have carried me throughout this academic journey. His guidance, provision, and love have been my constant source of strength from the very beginning until the completion of this doctoral milestone.

I would like to express my sincere appreciation to my supervisor, Dr. Majid Khan bin Majahar Ali, for his continuous support, expert guidance, and immense patience throughout my PhD journey. His vast knowledge and thoughtful feedback were instrumental in shaping this research, and his mentorship has left a lasting impact on both my academic and personal growth.

My heartfelt thanks go to my beloved husband, Firas, for his unconditional support, endless encouragement, and understanding. His presence has been a pillar of strength during the most challenging times. To my dear children, Yaman, Reeman, and Kayan, thank you for your patience, love, and for always being a source of joy and motivation. I am also deeply grateful to my family, especially my sister Fatimah, for her prayers, encouragement, and constant reassurance. Your support has meant the world to me. I would also like to thank my colleagues, friends, and everyone who stood by me throughout this journey—whether through academic collaboration, emotional support, or kind words when I needed them most.

This achievement is not mine alone. It is the result of the collective love, guidance, and encouragement I have been blessed with along the way. To each and every person who contributed, directly or indirectly, to this accomplishment, I offer my heartfelt gratitude.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
GPS	Global Positioning System
IoT	Internet of Thing
LASSO	Least Absolute Shrinkage and Selection Operator
MAPE	Mean absolute percentage error
MSE	Mean square error
PF	Precision Farming
R^2	Coefficient of Determination
SFTs	Smart Farming Technologies
SSE	Sum of square error
v-GHSD	v- Groove Hybrid Solar Drier
VIF	Variance Inflation Factor

LIST OF APPENDICES

PENGOPTIMUMAN KECEKAPAN PENGERINGAN RUMPAI LAUT MENGGUNAKAN SOLAR HIBRID DAN MODEL REGRESI TEGUH-JARANG

ABSTRAK

Analitik data dalam statistik amat penting untuk mengekstrak maklumat, mengenal pasti corak, dan membimbing keputusan. Dalam pertanian tepat, khususnya pengurusan pasca-tuai, cabaran timbul daripada kebergantungan kepada sensor IoT, kerumitan sistem, dan interaksi pemboleh ubah, yang membawa kepada isu seperti kepelbagaian, multikolineariti, dan kepekaan terhadap nilai terpencil. Menangani cabaran ini memerlukan peningkatan keterangkuman data, pengurusan data yang kukuh, dan kerjasama merentas sektor untuk membuka potensi penuh analitik. Kepelbagaian dalam sistem pertanian memberi kesan kepada hasil tanaman dan proses pasca-tuai. Heterogeniti dalam sensor, kaedah pengumpulan data, dan protokol penghantaran menyukarkan proses pengeringan pertanian. Multikolineariti, di mana pemboleh ubah bebas sangat berkorelasi, mencipta kesukaran dalam pemantauan pasca-tuai apabila data persekitaran yang bertindih daripada pelbagai sensor menutup kesan pemboleh ubah individu. Turun naik akibat perubahan persekitaran, kesilapan sensor, dan campur tangan manusia semakin merumitkan pemodelan, memerlukan kaedah statistik yang kukuh untuk menangani bunyi dan nilai terpencil. Set data yang dianalisis merangkumi 29 pemboleh ubah tunggal dan 406 pemboleh ubah interaksi, dengan jumlah keseluruhan 435 parameter apabila interaksi peringkat kedua diambil kira. Dengan hanya 1,914 pemerhatian bagi setiap pemboleh ubah, data berdimensi tinggi ini menekankan keperluan untuk pemilihan pemboleh ubah yang berkesan. Tujuh pemboleh ubah utama—T6, T7, T8, T11, H1, H5, dan PY—dikenal pasti sebagai penyumbang penting kepada kepelbagaian. Memberi tumpuan kepada

pemboleh ubah berkedudukan tinggi ini adalah penting untuk membangunkan model ramalan yang boleh dipercayai. Adaptive LASSO menunjukkan prestasi unggul dalam menangani heterogeniti dan multikolineariti, membentuk asas untuk analitik ramalan yang kukuh. Namun, nilai terpencil dan bunyi mengurangkan kebolehpercayaan model, memerlukan kaedah kukuh seperti MM, MM Hampel, dan MM Huber. Teknik ini mengurangkan kesan anomali, meningkatkan ketepatan, dan menambah baik konsistensi merentas model, menjadikannya penting untuk sistem pemantauan pasca-tuai yang boleh dipercayai. Model hibrid yang menggabungkan Adaptive LASSO dengan teknik regresi kukuh berjaya menangani cabaran multikolineariti dan nilai terpencil. Carta kawalan mengesahkan bahawa Adaptive LASSO dengan kaedah MM mencapai keseimbangan terbaik antara ketepatan dan kesesuaian model di bawah heterogeniti yang diubah suai. Penyelidikan masa depan boleh mengintegrasikan kaedah kukuh dengan teknik pembelajaran mesin dan pembelajaran mendalam yang maju, bersama analitik masa nyata, untuk mengurus set data berdimensi tinggi, meningkatkan pembuatan keputusan, dan menambah baik kecekapan dalam sistem pertanian tepat.

OPTIMIZATION SEAWEED DRYING EFFICIENCY USING HYBRID SOLAR DRYERS AND SPARSE ROBUST REGRESSION MODELS

ABSTRACT

Data analytics in statistics is vital for extracting insights, identifying patterns, and guiding decisions. In precision farming, particularly post-harvest management, challenges arise from IoT sensor dependency, system complexity, and variable interactions, leading to issues like variability, multicollinearity, and sensitivity to outliers. Addressing these challenges requires improved data inclusivity, robust data management, and cross-sector collaboration to unlock the full potential of analytics. Variability in agricultural systems impacts crop yield and post-harvest processes. Heterogeneity in sensors, data collection methods, and transmission protocols complicates agricultural drying. Multicollinearity, where independent variables are highly correlated, creates difficulties in post-harvest monitoring as overlapping environmental data from multiple sensors obscures the impact of individual variables. Fluctuations due to environmental changes, sensor errors, and human interventions further complicate modeling, requiring robust statistical methods capable of handling noise and outliers. The dataset analyzed includes 29 single variables and 406 interaction variables, totaling 435 parameters when second-order interactions are considered. With only 1,914 observations per variable, this high-dimensional data underscores the need for effective variable selection. Seven key variables—T6, T7, T8, T11, H1, H5, and PY—were identified as significant contributors to variability. Focusing on these high-ranking variables is critical for developing reliable predictive models. Adaptive LASSO demonstrated superior performance in handling heterogeneity and multicollinearity, forming the foundation for robust predictive analytics. However, outliers and noise reduce model reliability, necessitating robust

methods like MM, MM Hampel, and MM Huber. These techniques mitigate the impact of anomalies, improve accuracy, and enhance consistency across models, making them essential for reliable post-harvest monitoring systems. A hybrid model combining Adaptive LASSO with robust regression techniques effectively addressed challenges of multicollinearity and outliers. Control charts confirmed that Adaptive LASSO with MM methods achieved the best balance of accuracy and model fit under modified heterogeneity. Future research could integrate robust methods with advanced machine learning and deep learning frameworks, along with real-time analytics, to manage high-dimensional datasets, improve decision-making, and enhance efficiency in precision farming systems.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Data analytics in statistics is a powerful tool for interpreting complex datasets and deriving actionable insights, with visualization techniques playing a crucial role in summarizing data, identifying patterns, and informing decisions across various fields. This process involves a combination of statistical methods, techniques in data mining and visualization tools are used to convert raw data into valuable and actionable insights. The integration of these elements is essential for effective decision-making and strategic planning in diverse sectors (Kumari et al., 2024).

Data analytics significantly transforms agriculture and aquaculture by enabling more informed decision-making, optimizing operations, and promoting sustainable practices. This progress is fueled by the incorporation of advanced technologies like big data management, the Internet of Things (IoT), and machine learning. These technologies collectively enhance precision farming, improve resource management, and contribute to environmental sustainability (Raji et al., 2024).

Precision farming, also known as precision agriculture, is a modern farming management concept that leverages cutting-edge technologies such as the Global Positioning System (GPS), Geographic Information System (GIS), Internet of Things (IoT), Artificial Intelligence (AI), and machine learning to enhance optimization. field-level management, enhancing productivity, efficiency, and sustainability by addressing variability within fields and tailoring resource inputs to the specific needs of crops and livestock. This approach integrates information technology with traditional farming practices, revolutionizing agricultural production and offering significant economic and environmental benefits. By leveraging data collection and analysis technologies, farmers can map fields, monitor yield variations, and apply inputs like fertilizers and pesticides variably across different field zones, thereby improving resource use efficiency and minimizing environmental impact (Singh et al., 2021; Buick, 1997). Remote sensing and smart sensors play a vital role in monitoring crop health and environmental conditions, delivering real-time data to facilitate precise resource allocation (Ovchinnikov & Afanasyeva, 2023).

In dairy farming, precision technologies measure the physiological and production indicators of individual animals, allowing for better management strategies and early disease detection (Tutkun, 2023). Economically, precision farming benefits small-scale farmers by optimizing resource allocation, reducing costs, and increasing profitability through targeted interventions based on accurate data (Mahanto et al., 2024). However, the adoption of precision farming faces challenges, including significant investment in technology and infrastructure, which can be a barrier for farmers in developing regions (Rimpika et al., 2023). Additionally, the commodification of agricultural information and reliance on digital data may reduce farmers' reliance on traditional knowledge and public sector research, potentially impacting local agricultural practices (Wolf & Wood, 2010). While precision livestock farming improves productivity and animal welfare, it cannot fully replace the farmer's knowledge and experience, highlighting the need for a balanced approach (Cox, 2003). As precision farming technologies continue to evolve, it is essential to ensure they are accessible to all farmers and that their implementation supports sustainable and equitable agricultural development, balancing technological advancement with the preservation of local knowledge and practices.

The synergy between precision farming and data analytics is fundamentally transforming agriculture by integrating advanced technologies and data-driven insights to enhance productivity, sustainability, and resilience.

Data analytics is at the core of this process, analyzing large volumes of data from various sources, including satellite imagery and soil sensors, to generate actionable insights for better crop management (Chen et al., 2024). The precision enabled by these technologies supports sustainable farming by tailoring inputs to the specific needs of different field zones, thereby minimizing waste and promoting environmental stewardship (Roy et al., 2024). However, challenges such as data quality, the need for technical expertise, and the digital divide, particularly in developing regions, must be addressed through investments in data infrastructure, talent development, and policy support to fully realize the potential of precision farming. Overcoming these barriers will ensure that precision agriculture contributes not only to enhanced productivity and sustainability but also to global food security by making agriculture more resilient to climate variability (Chen et al., 2024)

In agriculture, the complexity and multivariate nature of data demand robust management systems that facilitate low-latency, high-efficiency computations, essential for applications such as weather forecasting, land usage analysis, and crop price prediction, which are pivotal in smart agriculture and agro-advisory systems (Gandhi, 2024). One important agriculture especially in Asia is seaweeds, which are divided into green (Chlorophyta), brown (Phaeophyta), and red algae (Rhodophyta), these are important marine organisms that contribute significantly to marine ecosystems by performing as primary producers and preserving coastal biodiversity (Ghaliaoui et al., 2024).

Seaweed, a diverse group of marine macro-algae, is increasingly recognized for its multifaceted benefits across food, health, industry, and environmental sustainability. Its rich nutritional profile, ecological advantages, and potential for sustainable development make it a valuable resource. Seaweed's applications range from being a nutritious food source to a sustainable alternative in various industries, contributing to global health and environmental goals. Seaweeds are high in critical elements including iodine, vitamins, proteins, and fiber, all of which improve thyroid function and general health (Dissanayake et al., 2024). Seaweeds are commonly used in the food, pharmaceutical, and cosmetic sectors because of their developing, expanding, and mixing features, which enhance product texture, shelf life, and nutritional profile (Sharma et al., 2023). Seaweed farming is a developing sector with significant environmental and economic advantages, with important phases including cultivation, harvesting, post-harvest processing, and product development. Seaweed is grown in marine habitats without the use of water or fertilizers. Environmental variables such as water temperature, salinity, and pH are monitored to maximize growth conditions (Izharuddin et al., 2023). Harvesting techniques, whether traditional or mechanical, are selected depending on size, and time is crucial to achieve ideal biochemical composition, such as protein concentration (Stedt et al., 2022).

Post-harvest management is critical in preserving seaweed quality, with advances in drying technology and cleaning methods aimed at improving product results (Ali et al., 2017). Seaweed-based product development, ranging from food to medicines, focuses on extracting bioactive substances such as polysaccharides while maintaining safety via complex drying processes (Zhu et al., 2021). Market measures, such as selling time and availability, have an impact on economic results, but post-harvest management and sustainable practices continue to be challenging (Sari et al., 2023). Addressing these difficulties via technical and local knowledge integration is critical to the industry's development and survival. In aquaculture, data analytics is crucial for environmental risk assessments, as demonstrated by studies on the sensitivity of seaweed across different life stages, emphasizing the need for comprehensive data to set accurate toxicity thresholds (Marteinson et al., 2024). Moreover, remote sensing and big data mining have uncovered correlations between aquaculture practices and environmental phenomena like seaweed, highlighting the delicate balance required between marine resource development and environmental protection (Liang & Cui, 2024).

Technological advancements, such as tray solar dryers, have exhibited improved drying kinetics and decreased contamination risk in Eucheuma cottonii (Amir et al., 2024). Seaweed drying is an important post-harvest process that has a substantial influence on its quality, drying velocity, and effectiveness. Various models and approaches have been developed to enhance this process. Solar and biomass drying technologies, which often use hybrid systems, provide higher drying rates and energy efficiency, with hybrid systems achieving a drying rate of 0.0893 g/s and a 13.36% efficiency (Mulyadi et al., 2018). Microwave drying has proven to be a fast and efficient alternative, achieving a drying rate of 30.29 g/min and an energy efficiency of 58.45%, making it well-suited for large-scale industrial applications (Hakim et al., 2020). Furthermore, non-hygroscopic drying models, especially for materials, provide reliable forecasts of drying rates by accounting for moisture fluctuation (Andrianantenaina et al., 2016).

Technological developments such as the v-Groove Hybrid Solar Drier (v-GHSD) and offshore drying solutions have shown promise in efficiently lowering moisture content, with offshore systems reducing moisture to 44.98% in 24 hours (Ali et al., 2015; Mayol et al., 2019). While each approach has benefits, concerns for energy consumption, environmental effects, and cost must be weighed, with solar and biomass processes providing more sustainable alternatives, but at slower drying rates than microwave technology.

The moisture content is a crucial factor in many sectors that affects the quality, preservation, and physical characteristics of materials. Multiple sophisticated techniques and technologies may be used to precisely test and regulate this parameter. In addition, high-frequency medium heating systems enable accurate regulation of moisture content by calibrating heating currents, therefore obviating the requirement for supplementary sensors and providing a cost-efficient solution (Yongtong et al., 2019).

Regulation of moisture content is crucial in the food sector to ensure product quality, preservation, and uniformity in manufacturing. In material industries, moisture levels have a significant impact on characteristics such as electrical conductivity and refractive index (Nielsen, 2010; Rani et al., 2015). Despite these gains, difficulties remain in guaranteeing precision and dependability across different materials and environments, requiring continuous study and technical enhancements to maximize moisture measuring and management techniques.

Optimizing crop yields, assuring quality, and sustaining the health of agricultural products need a comprehensive understanding and effective management of moisture content in agriculture. Various sophisticated methods and technologies have been developed to monitor and forecast moisture levels in varied settings. Wireless sensor networks offer an advanced method for monitoring soil moisture, allowing for the forecast of moisture levels and the effective management of irrigation, therefore improving the sustainability and efficiency of agricultural activities (Wen et al., 2017).

The interplay between post-harvest activities and agriculture is of utmost importance in assessing the quality, marketability, and economic worth of agricultural products. Various procedures such as transportation, processing, and storage have substantial impacts on minimizing losses and improving food security. Alawa and Bishie-Unung (2024) have demonstrated that implementing efficient post-harvest management practices may enhance market accessibility and mitigate the issue of farmers selling their crops at unfavourable prices, as exemplified by the case of cassava in Nigeria. The use of post-harvest technologies, which are impacted by variables like education, income, and extension services, is crucial for mitigating spoilage and losses. This has been shown in Kenya with maize and mangoes (Onkware et al., 2021). Recent technological progress, such as remote sensing and sensor technologies, has significantly transformed post-harvest management by allowing immediate monitoring and improved decision-making to maintain grain quality during storage (Rodrigues et al., 2024).

Despite occasional hesitancy caused by market volatility, small-scale farmers can get advantages from post-harvest storage by carefully scheduling their sales to take advantage of price hikes during off-season months (Priya & Mitra 2020). Notwithstanding the difficulties, the use of sophisticated technology and sustainable methods in post-harvest operations, backed by suitable legislation, infrastructure, and farmer education, is crucial for reducing losses and improving food security.

Drying is a critical post-harvest process for preserving food and biological products such as seaweed, yet traditional open-air drying methods often suffer from inconsistent results due to uncontrolled environmental factors like temperature, humidity, and weather variability. These fluctuations can significantly affect product quality and energy efficiency. To address these limitations, this study proposes the use of a hybrid solar dryer, which converts solar energy into electricity to support a controlled drying environment. By integrating sensors that monitor and regulate key variables—such as temperature, humidity, solar radiation, and airflow—the system ensures optimal drying conditions. However, one of the key challenges in implementing such systems lies in the variability of sensor data, as different sensors operate within vastly different value ranges. This heterogeneity can introduce multicollinearity and outliers into the dataset, distorting predictive analysis. Moreover, there is a notable lack of research exploring the impact of high-ranking variables across large sensor networks—particularly in the context of seaweed drying—despite the availability of extensive sensor data (up to 435 variables). Existing models often overlook the combined effects of these variables or fail to address data quality issues arising from model complexity. This study aims to fill these gaps by developing a robust framework that integrates advanced modeling techniques, including sparse and robust regression, to manage data variability and enhance prediction accuracy. Ultimately, this research contributes a novel approach for improving the efficiency, reliability, and scalability of solar-powered drying systems in controlled environments.

1.2 Seaweed in Malaysia

Seaweed farming in Malaysia has emerged as a crucial sector with significant potential for economic development and environmental sustainability, particularly through the cultivation of species that are valuable for biofuel production, nutritional benefits, and export, notably in carrageenan production (Nor et al., 2020; Olanrewaju et al., 2024).

Over the past decade, Malaysia’s carrageenan industry has experienced significant changes, particularly in seaweed cultivation and processing capacity. In 2014, the country produced over 2,000 tonnes of dried seaweed monthly, though only 10–50% of this output was utilized by domestic processors, with the remainder exported or sold via intermediaries (Neish et al., 2019). That same year, carrageenan exports from Malaysia accounted for just over 1,000 tonnes—approximately 2% of the global market. By 2016, production had peaked at 205,989 tonnes (wet weight), valued at approximately USD 24.83 million (Hurtado et al., 2018). However, this upward trend did not persist; production sharply declined to an estimated 30,000 tonnes by 2021, largely due to environmental stressors, disease outbreaks, and competition from neighboring countries such as Indonesia and the Philippines (Ali et al., 2020). In response to these challenges, M.K.M. Ali has played a crucial role in developing innovative techniques aimed at improving both the cultivation and post-harvest handling of seaweed. His research has focused on micropropagation and sea-based nurseries for Kappaphycus species, as well as solar drying innovations such as v-groove hybrid solar dryers, all designed to enhance product quality and processing efficiency (Ali & Ismail, 2017; Ali et al., 2014). Moreover, his application of response surface methodology (RSM) has optimized the extraction conditions for refined carrageenan from Kappaphycus striatum, significantly improving both yield and purity (Ali et al., 2014). These scientific advancements are essential to restoring Malaysia's position in the global carrageenan supply chain and ensuring the industry’s long-term sustainability.

The industry supports the livelihoods of coastal communities, especially in regions like Sabah, by providing income and enhancing food security, thanks to the nutritional richness of seaweeds, which are high in dietary fiber, minerals, and essential amino acids (Chin et al., 2023; Ali et al., 2020). However, the sector has faced challenges such as declining production since 2013, primarily due to socio-economic and environmental issues like ice-ice syndrome and pest outbreaks, which have negatively impacted crop quality and yield (Asri et al., 2021; Kambey et al., 2021).

The Malaysian government's initiatives under the National Key Economic Area aim to promote seaweed farming as a sustainable livelihood option, with projects like the Seaweed Cluster Project working to optimize farming practices and improve quality, although these efforts are hindered by poor stakeholder participation and complex market dynamics (Ali et al., 2020; Nor et al., 2017). Technological advancements, including micropropagation and the use of biostimulants, are being explored to enhance productivity and sustainability (Ali et al., 2020). Addressing these challenges, particularly by strengthening local cooperatives, improving governance, and supporting the role of migrant workers, is essential for realizing the full potential of the Malaysian seaweed industry and ensuring its long-term resilience and sustainability.

The processing of seaweed encompasses a variety of advanced methods and technologies aimed at enhancing its utility across food, feed, and industrial applications by improving efficiency, reducing contaminants, and extracting valuable compounds. Transforming batch processes into continuous production lines has significantly increased the output of seasoned seaweed products while reducing costs and environmental impact (Mongkolkittaveepol et al., 2023). Effective contaminant reduction in seaweed, particularly iodine and heavy metals, can be achieved through washing, blanching, rehydrating, and heat treatments, which significantly lower elemental concentrations (Van Tuinen et al., 2023).

Drying seaweed for energy production optimizes it for use as a renewable fuel, demonstrated by reducing moisture content to 12% for gasification (Nazemi et al., 2022). An integrated biorefinery approach for brown seaweed maximizes resource utilization by extracting multiple products, such as alginic acid and protein concentrate, while managing effluents sustainably (Baghel et al., 2020). Additionally, producing high-strength seaweed extract fibers through ultrasonic treatment offers potential applications in the textile industry (Liqing & Ke, 2019). Despite these advancements, ongoing challenges include balancing process efficiency with environmental sustainability, requiring further research and innovation to fully harness seaweed's potential as a sustainable resource.

1.3 Problem Statement

Data analytics, while a powerful tool for extracting insights from large datasets, faces several significant gaps that hinder its effectiveness. These gaps arise from various challenges, including data availability, quality, and representation, particularly concerning marginalized groups. Additionally, the technical and ethical challenges associated with big data analytics further complicate the landscape.

Data analytics faces several critical challenges that limit its effectiveness, especially in the context of marginalized populations and the technical and ethical challenges associated with big data. A significant gap arises from the underrepresentation and invisibility of marginalized groups, such as LGBTIQ+ individuals, within data archives, leading to biased insights and perpetuating mainstream perspectives. The rapid growth of big data further complicates this landscape, with issues related to managing its volume, variety, and velocity, making data engineering and cleaning processes increasingly complex. Conventional data management tools often fall short in handling the dynamic nature of big data, requiring new solutions to ensure efficient analysis. Additionally, the quality and integrity of data, especially in sectors like healthcare, directly impact decision-making, with poor data quality potentially leading to adverse outcomes. Ethical considerations, including responsible data use, model accuracy, and regulatory compliance, pose further challenges, particularly in sensitive areas such as finance and healthcare. Furthermore, inconsistencies in data formats across organizations and sectors hinder collaboration and data-sharing efforts, particularly in areas like mobility and smart city initiatives, where standardization is crucial. Addressing these challenges requires improving data inclusivity, advancing data management techniques, and fostering cross-sectoral collaboration to realize the full potential of data analytics.

The classification and prediction of significant variables affecting seaweed farming present considerable challenges, especially in regions that rely heavily on uncontrolled environmental conditions such as open-sea aquaculture. Unlike more advanced countries like Singapore and Malaysia, which have adopted controlled environments through building-based or indoor vertical farming systems, many developing regions still depend on natural factors like temperature, salinity, and nutrient flow. These uncontrollable elements can significantly affect the growth rate, nutritional content, and overall yield of seaweed. Inconsistent conditions often result in poor biomass quality, reduced productivity, and vulnerability to diseases or environmental stress. This issue extends to post-harvest processes as well—particularly drying—where traditional methods in open environments suffer from external factors like weather, temperature fluctuations, and humidity. Such variables can lead to uneven drying, microbial contamination, and loss of essential bioactive compounds. In contrast, controlled environments equipped with smart sensors allow for precise monitoring and regulation of temperature, humidity, and airflow, ensuring better drying efficiency and preserving the quality of the harvested seaweed.

From a data analysis perspective, the presence of ultra-dimensional data—with numerous environmental and biological parameters—introduces analytic complexity. Multicollinearity among variables and the presence of outliers often lead to unstable models and unreliable predictions. These modeling limitations make it harder to optimize seaweed farming processes or forecast outcomes like yield, quality, or growth cycles. Despite the growing availability of machine learning and advanced statistical tools, current research remains limited in identifying which variables have the strongest predictive influence and understanding their direct impact on outcomes. Additionally, there is a gap in analyzing prediction errors before and after data variation, especially using high-ranking variables within hybrid models or robust regression techniques. This lack of insight ultimately hinders the development of adaptive, data-driven farming systems that can respond intelligently to environmental fluctuations, which is essential for scaling sustainable seaweed production.

The variability of the technologies and data sources involved poses a significant difficulty in post-harvest monitoring. Heterogeneity refers to the diversity and variations in sensor kinds, data-gathering techniques, and transmission protocols. Sensors detecting critical variables such as temperature, humidity, or moisture content, for example, may be manufactured by multiple companies, each with its calibration standards, sensitivities, and operational ranges. Furthermore, these sensors may be placed in a variety of places with varied circumstances, exacerbating discrepancies in the data they acquire.

This heterogeneity puts noise into the data, making it difficult to obtain consistent and trustworthy measurements from several sensors and locations. When such inconsistent data is incorporated into predictive models, particularly ones that rely heavily on important factors such as temperature and humidity, their sensitivity can be severely reduced. Model sensitivity refers to the model's capacity to effectively adapt to changes in input variables and anticipate outcomes based on those changes. If the input data is noisy or inconsistent, the model may miss tiny but significant changes that might signal rotting or quality decline in agricultural goods. This failure can lead to incorrect projections and poor decision-making, resulting in financial losses and lower product quality. Recent research highlights how sensor variability might reduce model sensitivity and overall dependability of post-harvest monitoring systems.

Another key challenge in post-harvest monitoring systems is multicollinearity as inter-variable interactions. Multicollinearity arises when two or more independent variables in a predictive model are strongly linked, making it difficult to isolate each variable's influence on the dependent result. In agriculture, this problem frequently emerges when many sensors in various locations collect relevant environmental factors such as temperature, moisture content, radiation, and humidity. For sensors that measure these variables, there are 29 single variables (T1 to T29), (H1, H5) and PY and 406 interaction variables due to the large number of sensors, variable selection becomes highly important. for example, may show substantial correlations due to comparable environmental influences. Furthermore, the physical placement of sensors might cause data changes, depending on whether they are in shaded regions, exposed to direct sunlight, or located in zones with variable airflow patterns. These variances might cause interaction effects, which create complicated interactions between variables and complicate the model even further. When multicollinearity and interaction effects exist, the model's estimated coefficients become unstable, decreasing its capacity to reliably predict outcomes. This can lead to incorrect data interpretations since individual variables' genuine impacts are disguised by their connections with others. Addressing the difficulties of multicollinearity in agricultural data is critical, and improved solutions are required to reduce its influence and get more trustworthy model outputs.

There is a critical need for models that can endure the challenges posed by outliers in agricultural data. Robustness in statistical modeling refers to a model's capacity to perform effectively even when its underlying assumptions are violated or when the data includes outliers or noise. In agriculture, data often fluctuates due to various factors, such as environmental changes, sensor errors, and human interventions. For instance, unexpected weather conditions or equipment failures can introduce anomalies that could lead to inaccurate predictions if the model lacks robustness. Developing robust models is vital to ensuring that post-harvest monitoring systems remain reliable and accurate over time. These models are less sensitive to irregularities in the data, allowing them to generate accurate predictions even when the data is imperfect. This is especially important in post-harvest management, where decisions based on unreliable models could result in significant financial losses, whether through wasted resources or reduced product quality.

To effectively tackle the challenges of multicollinearity and the need for robustness, there is an increasing demand for optimized hybrid models. These models combine the strengths of different modeling techniques to manage multicollinearity while ensuring robustness. For example, traditional statistical methods like ridge regression, which are specifically designed to address multicollinearity, can be integrated with machine learning algorithms known for their robustness and adaptability to complex datasets. Optimizing such a hybrid model involves carefully balancing various factors, including model complexity, interpretability, computational efficiency, and prediction accuracy. The aim is to create a model that not only mitigates the effects of multicollinearity but also remains robust in the face of data variability. This process requires advanced model selection techniques, such as cross-validation, where the model is evaluated on different subsets of data to ensure consistent performance across various scenarios. Additionally, hyperparameter tuning is often necessary to fine-tune the model's performance and achieve an optimal balance between accuracy and robustness, leading to more reliable and efficient post-harvest monitoring systems.

A final challenge emerges from traditional econometric and statistical practices, which often advocate for the reintroduction of individual parameters into models if they are found to be statistically significant. This practice is typically aimed at enhancing model interpretability and capturing the marginal effects of specific variables. Indeed, adding a single variable—such as average temperature, soil pH, or fertilizer input—can offer certain benefits. It may improve the model’s explanatory power, yield more interpretable coefficients for domain experts, and highlight direct associations between that variable and the outcome of interest. When the variable has strong theoretical justification, its inclusion can support practical decision-making and policy development. However, in complex systems like agricultural monitoring, this approach can also be problematic. Adding a single parameter back into a model without considering the broader context—such as potential interaction effects or the presence of multicollinearity—can result in oversimplification. This may cause the model to overlook the systemic nature of agricultural dynamics, where outcomes are shaped by the combined effects of multiple interacting variables. For instance, reintroducing temperature alone might ignore how it interacts with humidity, soil moisture, or irrigation practices in subtle yet crucial ways. Such interactions are essential in agriculture, where environmental factors often have compounding or conditional effects. The impact on model performance can also be mixed. While the variable might temporarily improve statistical fit (e.g., R^2), it can introduce instability if it is highly correlated with other predictors. This multicollinearity can inflate standard errors, distort coefficient estimates, and reduce the model's reliability across different contexts. The variable's apparent significance may stem from shared variance with others, leading to misleading conclusions. To mitigate these risks, researchers should consider more holistic modeling approaches—such as interaction-aware regression, generalized additive models, or machine learning methods—that can accommodate non-linear relationships and complex variable interdependencies.

1.4 Objectives of the Study

This work's objectives are to:

i. To identify significant variables that affect heterogeneity using boxplot and VIF.

ii. To identify significant parameters after heterogeneity that directly influence using high-ranking variables (50, 100, 150, 200, 250, and 300).

iii. To optimize the most important variables across three spares regression: LASSO, Elastic Net, and adaptive LASSO using high-ranking variables with minimized MAPE, MSE, SSE and high R^2.

iv. To identify outliers for before and after heterogeneity for high-ranking variables using Sparse regression and robust (M, MM, S, MM-Hampel, MM-Huber, MM-Bi square, M-Huber, M-Hampel, M-Tukey) model.

v. To compare the developed hybrid model after heterogeneity and adding back removed single parameters to understand the multicollinearity and outlier's effects using a control chart of residual errors.

1.5 Research Questions

Based on the objectives of the study, the following research questions are proposed:

i. Research Question 1: What are the key variables contributing to data heterogeneity in the drying process of seaweed using the v-Groove Hybrid Solar Dryer (v-GHSD)?

ii. Research Question 2: Which high-ranking environmental variables significantly influence seaweed drying efficiency after accounting for heterogeneity?

iii. Research Question 3: Which sparse regression model among LASSO, Elastic Net, and Adaptive LASSO achieves the best predictive accuracy and variable selection for the seaweed drying process?

iv. Research Question 4: How effective are robust regression models in identifying outliers in high-dimensional environmental data before and after heterogeneity filtering?

v. Research Question 5: What is the effect of reintroducing previously excluded single parameters on model accuracy and residual distribution, particularly in the presence of multicollinearity and outliers?

1.6 Hypotheses

The study is guided by the following hypotheses:

i. Hypothesis 1: There is a statistically significant difference in variance among the environmental variables, indicating the presence of heterogeneity affecting drying performance.

ii. Hypothesis 2: A subset of high-ranking variables selected after heterogeneity adjustment has a significant impact on drying outcomes such as moisture reduction and energy efficiency.

iii. Hypothesis 3: Adaptive LASSO will outperform LASSO and Elastic Net in terms of prediction accuracy and model fit, as evidenced by minimized MAPE, MSE, SSE, and maximized R^2.

iv. Hypothesis 4: Robust regression models will detect statistically significant outliers that are overlooked by traditional sparse models, especially when heterogeneity is accounted for.

v. Hypothesis 5: Reintegrating statistically significant single variables into the hybrid model improves performance metrics and stabilizes residual error patterns without significantly increasing multicollinearity.

1.7 Scope and Limitation

The objective of this study is to analyze the drying process of seaweed using a V-Groove Hybrid Sun Dryer (v-GHSD), as illustrated in Figure 3.2, with an emphasis on the critical environment and operational variables such as temperature, ambient relative humidity, chamber relative humidity, and sun radiation. Data was collected for 29 important factors, each with 1,914 data points, to provide a thorough basis for examining the correlations between these variables and drying efficiency. However, the study's scope is limited by the specific environmental conditions under which the data was collected, which may not fully represent all potential scenarios in which the drier could be utilized. Furthermore, differences in sensor precision and positioning, as well as the emphasis on a certain variety of seaweed, may generate variations and limit the applicability of the findings. These limitations highlight the need for more studies into different environmental variables, sensor systems, and varieties of seaweed to increase the application of the results.

The analysis is based on the assumption that certain statistical conditions are met, including normality, independence, homoscedasticity, linearity, randomness, and the absence of multicollinearity among the variables. These assumptions are critical to providing the validity and accuracy of the conclusions. However, the study's challenges include the possibility of rejecting these assumptions, which might result in biased or inaccurate outcomes. Additionally, the study is constrained by the specific environmental conditions during data collection, which may not represent all possible scenarios where the drier could be used. These limitations suggest that the findings may not fully capture the complexities of the drying process, indicating a need for further research to validate the results under different conditions and ensure the robustness of the conclusions.

This study employs machine learning models, including Lasso, Elastic Net, and Adaptive Lasso, to analyze and predict the drying process of seaweed using a v-Groove Hybrid Solar Drier (v-GHSD). These models are particularly suited for handling complex data structures and addressing multicollinearity issues that arise when working with environmental factors such as temperature, ambient relative humidity, chamber relative humidity, and solar radiation. Ridge regression is used to tackle multicollinearity by applying a penalty that shrinks the coefficients of correlated variables, enhancing model stability and prediction accuracy. LASSO regression builds on this by introducing a regularization technique that not only shrinks coefficients but also forces some to be zero, effectively selecting the most significant variables and simplifying the model. Elastic Net combines the strengths of both Ridge and LASSO, striking a balance between variable selection and handling multicollinearity, making it ideal for datasets with highly correlated predictors. Adaptive Lasso further refines this process by applying varying penalties to different coefficients, allowing for more precise variable selection and improved model accuracy. By leveraging these machine learning models, the study seeks to uncover the complex relationships between environmental factors and drying efficiency, developing a predictive framework that is both robust and interpretable. The use of these advanced methods enhances the reliability of the results and provides valuable insights for optimizing the drying process, with potential implications for improving the operational efficiency of the v-GHSD.

1.8 Significance of the study

The agriculture sector is a cornerstone of Malaysia's economy, and it has recently experienced significant advancements due to the integration of IoT sensors and data analytics. These innovations have revolutionized the way data is collected and analyzed, particularly in areas like seaweed drying. IoT sensors play a crucial role in tracking various drying parameters in real time, delivering accurate data on parameters such as temperature, humidity, and moisture content. By leveraging this sensor data through advanced data analytics and machine learning techniques, the proposed model can significantly enhance the effectiveness of the seaweed drying method, directly benefiting farmers and industries by enhancing productivity and product quality. This research identifies the key factors influencing both moisture content removal and collector efficiency. A notable feature of this study is its focus on interaction factors within the data, ensuring that all potential models are thoroughly considered. The four-phase approach adopted in this project has resulted in an efficient predictive model, now capable of accurately forecasting the parameters related to moisture content removal and collector efficiency using a solar drier. The developed Standard Operating Procedure (SOP) is adaptable across different models, Offering an effective approach to address the challenges of big data modeling in agriculture by emphasizing critical parameters derived from IoT sensor data.

1.9 Thesis Framework

This thesis is structured into five chapters, each designed to systematically guide the reader through the research process. Chapter 1 introduces the study by outlining the background and context of the research problem, clearly stating the problem statement, research objectives, and the key questions the study aims to answer. It also defines the scope of the research, along with its limitations and any assumptions made. Chapter 2 provides a comprehensive review of the relevant literature, examining previous studies, theoretical frameworks, and key concepts related to the topic. This review not only highlights existing knowledge but also identifies gaps that the current study aims to address. Chapter 3 discusses the research methodology, detailing the research design, data collection methods, and analytical techniques employed. It also includes justifications for the chosen methods, as well as the reliability and validity of the data. Chapter 4 presents and analyzes the research findings, using visual aids such as tables and charts where appropriate. This chapter also includes a discussion of the results in the context of the existing literature, drawing comparisons and offering interpretations. Finally, Chapter 5 concludes the thesis by summarizing the main findings, drawing overall conclusions. It also suggests directions for future research based on the outcomes and limitations of the study.

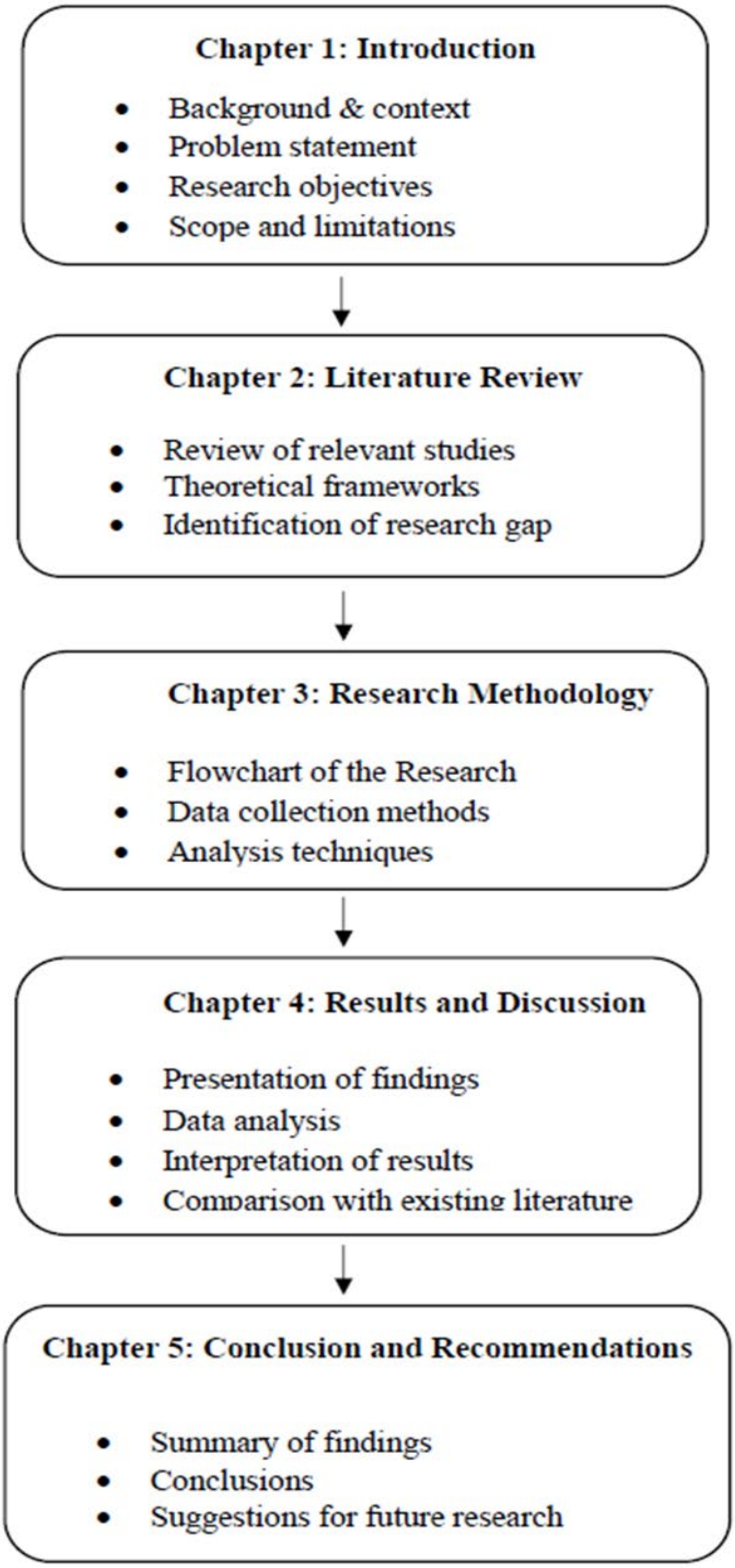

Figure 1.1 Thesis Framework