Automated Tuna Freshness Assessment via Gas Sensors and Machine Learning Algorithms

1st Nyoman Raflly Pratama
School of Electrical Engineering
Telkom University
Bandung, Indonesia
nyomanraflly@student.telkomuniversity.ac.id

2nd Ledya Novamizanti
School of Electrical Engineering
Telkom University
Bandung, Indonesia
ledyaldn@telkomuniversity.ac.id

3rd Dedy Rahman Wijaya
School of Appllied Science
Telkom University
Bandung, Indonesia
dedyrw@telkomuniversity.ac.id

Abstract— **Ensuring the safety and health of fish products is crucial for public health, with tuna being Indonesia's second most popular fishery product. Tuna freshness is a key indicator of seafood safety, directly impacting both nutritional quality and contamination risk. This study compares the K-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machine (SVM) algorithms to assess and classify tuna freshness, offering an accurate and efficient approach. A machine learning model categorized Tuna freshness based on the gases emitted, utilizing a dataset of 58,389 records. Gas changes were detected using the MQ-135, MQ-9, and MQ-2 sensors, which are highly sensitive to gases like ammonia, methane, and alcohol, commonly associated with spoilage. The KNN, Naive Bayes, and SVM algorithms were then applied to classify the sensor data. KNN and SVM achieved an accuracy of 99%, while Naive Bayes reached 90%. The high accuracy of these methods highlights their potential as practical tools for the fishing industry, enabling suppliers and retailers to assess tuna freshness more effectively. This method could significantly improve consumer safety by ensuring only high-quality, fresh products reach the market. Additionally, automation offers substantial time savings, facilitates faster decision-making, and reduces reliance on manual inspections prone to human error.**

Keywords—tuna, classification, gas sensor, machine learning

I. INTRODUCTION

A food product derived from marine creatures is called seafood. Consuming it has several health benefits, mainly because of its high level of Omega-3 w-3 polyunsaturated fatty acids, including EPA and DHA, which are known to prevent and treat diseases [1]. When ingested in sufficient quantities, one noteworthy advantage is the prevention of premature births. Seafood's high protein, low bad fat, abundant mineral content, and low cholesterol are further advantageous qualities.

People in Indonesia have a strong preference for seafood. The Republic of Indonesia's Ministry of Marine Affairs and Fisheries (KKP) reported that per capita fish consumption was 56.48 kg [2]. The amount of fish consumed nationally per person climbed to 58.48 kg by October 2023 [2]. Tuna, being the second most popular fishery product in the nation, plays an essential role in the economy and consumer health, yet is notably perishable due to its neutral pH and unsaturated fatty acids. Consequently, tuna freshness deteriorates within 8-20 hours at room temperature, leading to both quality loss and potential health risks if consumed beyond safe limits [3].

Ensuring seafood safety is thus crucial, with the FAO overseeing regulations to maintain the quality standards of fish products globally [4], [5], [6]. However, traditional methods of freshness detection, such as visual inspection or chemical analysis, are time-consuming and often lack the accuracy needed for timely assessments, especially at scale.

This study addresses these limitations by proposing an electronic nose (e-nose) system combined with machine learning to assess tuna freshness more accurately and efficiently. Utilizing MQ-135, MQ-9, and MQ-2 gas sensors sensitive to spoilage-related compounds, this research leverages a dataset of 58,389 records. Machine learning models (KNN, SVM, and Naive Bayes) were applied, achieving 99% accuracy with KNN and SVM, surpassing the conventional inspection's reliability. By optimizing hyperparameters and filtering noise, this approach introduces a novel application of Machine Learning in the seafood industry, potentially transforming freshness detection and ensuring only high-quality tuna reaches consumers.

II. RELATED WORKS

The usage of E-Nose devices, such as the PEN3, TGS series, and MQ series, is one of the possible options based on the difficulties that have been given and various literature references [7], [8], [9]. Machine learning methods are used for the computation and classification of the detected seafood quality. Manual analysis is used to select the parameter settings.

Recent studies have explored the use of gas sensors to detect the freshness of seafood products. These studies have found that using gas sensors such as MQ series can reliably detect gas emissions from fish, providing essential data for assessing freshness. For example, recent research has highlighted the potential of these sensors in developing electronic nose (e-nose) systems that mimic human olfaction to determine food quality.

In the studies used as references [7], [8], [9], MQ series gas sensors are utilized to detect levels of alcohol, butane, propane, methane, LPG, hydrogen, carbon monoxide, ammonia, and hydrogen sulfide as indicators of spoilage in seafood products. Total Viable Count (TVC) is a measurement that estimates the total number of live microorganisms in a sample, including various species of bacteria, yeast, and fungi. E. Yavuzer [7] study state that the results and gas analyses detected using MQ3, MQ4, MQ5, MQ8, MQ9, and MQ135 sensors indicate that the compounds released from fish during the spoilage process increase over time. However, previous research often lacked robust classification accuracy, with some studies facing challenges in optimizing gas detection parameters for diverse environmental conditions and specific seafood types. For example, Yavuzer [7] used MQ sensors to measure gases like ammonia and hydrogen sulfide to determine fish freshness but did not incorporate advanced noise-filtering techniques, which could lead to lower accuracy in varying storage conditions.

This study addresses these limitations by integrating an e-nose system with Machine Learning (ML) algorithms, including k-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machine (SVM), for enhanced classification accuracy. By employing a dataset of 58,389 records and optimizing hyperparameters, this research mitigates the issues of environmental sensitivity and parameter inconsistency that have affected previous studies. Unlike previous research, this approach applies noise filtering and hyperparameter tuning, which contribute to a highly reliable freshness classification system, achieving a target accuracy of 99%.

The freshness of seafood products can be assessed based on several factors, such as changes in texture from soft to slightly fibrous, alterations in taste when consumed, and the odor detectable in crab meat [10]. Seafood typically begins to emit an unpleasant odor within 3-7 days in open air, which is a clear indication that it is no longer safe to eat. Detecting whether seafood is fresh or less fresh in real-time, before physical signs such as changes in texture, surface slime, and odor become apparent, is a key focus of this research.

III. MATERIALS AND METHOD

A. *Dataset Acquisition*

Object were taken from fresh market using box filled with ice cubes to maintain freshness [14]. Then instantly put into an airtight box with room temperature (25°C) and a laptop are coupled to a gas sensor circuit in this training equipment, which when triggered automatically detect the gas levels in the airtight box. A 10-hour sampling procedure is used for every sample in order to measure the amount of gas that the fish emit. The data obtained by the program was stored and then labeled according to the fish quality obtained from expert opinion. Four experts who work in multinational supermarkets in Indonesia are consulted regarding fish quality and quality control.

The device used for the experiment is shown in Fig. 1. It includes an ESP S3 Devkit-C microcontroller connected to a gas sensor array consisting of two MQ-135 sensors (one calibrated for carbon dioxide and the other for ammonia), one MQ-9 sensor, and one MQ-2 sensor [11], [12], [13], as detailed in Table I. The setup was placed in an airtight container serving as the sample chamber. Samples were immediately placed in the chamber for detection, as shown in Fig. 2. The sampling process took up to ten hours to collect data on the tuna transitioning from fresh to spoiled at room temperature (25°C), resulting in 58,389 records, which were stored in a CSV file format.

Fig. 1. Electric nose prototype

Fig. 2. Tuna sample inside chamber

TABLE I. FUNCTION OF GAS SENSOR

No.	Gas Sensor	Gas Detection
1.	MQ 135	Ammonia
2.	MQ 135	Carbon Diokside (CO_2).
3.	MQ 9	Methane
4.	MQ 2	Alcohol

The gathered information was carefully analyzed and utilized to develop machine learning models that could precisely determine the fish's freshness. The dataset was enhanced with trustworthy and reputable information by adding professional assessments for fish quality labeling. The machine learning models were further strengthened by this expert-driven labeling, which significantly increased the freshness predictions' overall reliability and accuracy.

B. *Proposed of Method*

The proposed method involves multiple phases of dataset processing, including hyperparameter optimization, training, testing, and evaluation, with the goal of achieving the highest level of accuracy. Fig. 3 illustrates the suggested approach.

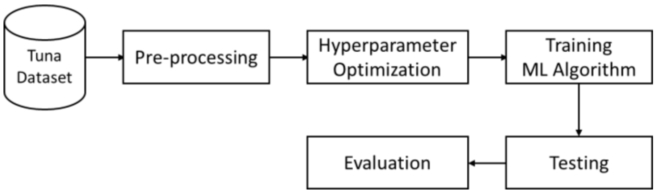

Fig. 3. Proposed of Method

The proposed approach for detecting the quality of tuna uses KNN, SVM, and Naive Bayes, with grid search applied for both regression and classification tasks. The KNN, SVM, and Naive Bayes algorithms were employed to optimize the hyperparameters, aiming to achieve optimal values for maximum classification accuracy and minimal error in the regression task. The parameters used in each method are displayed in Table II. By efficiently optimizing the hyperparameters, model performance was improved, reducing prediction error and increasing model accuracy.

TABLE II. HYPERPARAMETER OPTIMIZATION FOR TUNA CLASSIFICATION

Algorithm	Parameter	Value	Best parameter
KNN	n_neighbors	5	5
		10	
		20	
		50	
		100	
		150	
	weights	uniform	uniform
		distance	
	algorithm	auto	Brute
		ball_tree	
		kd_tree	
		brute	
SVM	C	0.01	100
		0.1	
		1	
		10	
		100	
	gamma	0.1	0.0001
		0.01	
		0.001	
		0.0001	
	kernel	rbf	rbf
		linear	
Naive Bayes	Var_smoothing	1e -9	1e -9
		1e -8	
		1e -7	
		1e -6	
		1e -5	
		1e -4	
		1e -3	
		1e -2	
		1e -1	
		1	
		10	
		100	

The raw data obtained from the gas sensors requires minimal preprocessing, To address this, Pareto scaling was applied. Pareto scaling normalizes the data by dividing each feature by the square root of its standard deviation, effectively reducing the impact of large outliers without entirely removing their influence. This method improves model sensitivity, particularly for algorithms like KNN and Naive Bayes, where feature scale can impact performance given the quality and consistency of the measurements. Equation (1) shows the formula for pareto scaling [15].

$$X'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{8j}} \quad (1)$$

This data is then split to couple sets: 80% of the data is allocated for training the machine learning models, while the rest of 20% is reserved for testing and validation purposes. The final 20% of the dataset, which was reserved for testing, is used to evaluate the performance of both models. During the machine learning training, the data collected from the e-nose gas sensor is saved in a CSV file format. This data is then utilized to train three algorithms, KNN, SVM and Naive Bayes. The process begins by separating the data into features. The freshness levels are categorized into three classes: fresh, less fresh, and not fresh. The evaluation metrics used in the evaluation are accuracy, precision, recall, F1 score, and confusion matrix.

IV. RESULTS AND DISCUSSION

Fig. 4 depicts a feature correlation matrix, which is a heatmap illustrating the correlations between four different feature: Ammonia (MQ-135), CO_2 (MQ-135), Alcohol (MQ-2), and Methane (MQ-9). The value range from 0 to 2, where 0 indicates no correlation and 1 indicates a perfect positive correlation. The heatmap uses a color gradient ranging from blue to red. Red represents strong positive correlations (closer to 1), while blue indicates moderate correlations (closer to 0.8).

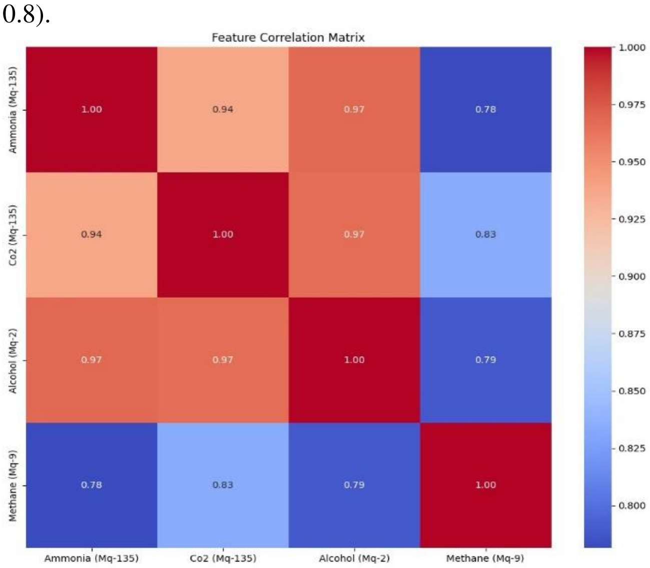

Fig. 4. Correlation Matrix

The gases Ammonia, CO_2, and Alcohol exhibit strong positive correlations, suggesting they may share common environmental factors or sources. This high level of correlation implies that fluctuations in one of these gases could be reflective of changes in the others. However, Methane (MQ-9) shows noticeably lower correlations with the other gases, particularly when compared to the strong correlations between Ammonia, CO_2, and Alcohol. This suggests that Methane may be influenced by different or independent variables, such as a sensor sensitivity variation.

Table III provides a summary of the hyperparameter optimization outcomes for the tuna freshness classification. KNN and SVM was able to classify the freshness of tuna with minimum errors, as seen by its high accuracy of 99%. In contrast, Naive Bayes achieved a lower accuracy of 90%, which, while still effective, indicates a greater degree of misclassification compared to KNN and SVM.

TABLE III. HYPERPARAMETER OPTIMIZATION CLASSIFICATION OF TUNA

No	Algorithm	Best Parameters	Accuracy
1.	KNN	N_neighbors : 5 Weights : uniform Algorithm : brute	99%
2.	Naive Bayes	Var_smoothing	90%
3.	SVM	C:100 Gamma : 0.0001 Kernel : rbf	99%

The performance for the KNN model can be seen in Fig. 5. With only 13 misclassifications and 11,665 correct classifications, the model achieved a 99% accuracy rate, demonstrating its high reliability in predicting tuna freshness. This impressive result highlights the effectiveness of the KNN algorithm in handling complex data, ensuring that fresh and spoiled tuna are accurately classified. In Fig. 6, the Naive Bayes algorithm's confusion matrix classification of tuna freshness is showcased. It recorded 1,146 misclassifications and 10,532 correct classifications, resulting in a 90% accuracy rate, indicating that it performs well but is less reliable compared to KNN and SVM. In Fig. 7, the SVM algorithm's confusion matrix shows only 9 misclassifications and 11,669 correct classifications, achieving a 99% accuracy rate, which is slightly better than KNN.

Showcasing the model's excellent reliability and robustness in distinguishing between fresh and spoiled tuna with fewer errors. The consistent performance of SVM and KNN at the 99% accuracy mark underscores their suitability for quality control processes, where precise identification of spoilage is crucial. The lower accuracy of Naive Bayes suggests that it might be more sensitive to data variance, making it less ideal for applications requiring high precision. Additionally, Naive Bayes may not handle the complex relationships between variables as effectively as KNN and SVM, which are better suited to the classification task in this specific context. This makes SVM and KNN more preferable choices for use in tuna freshness prediction systems.

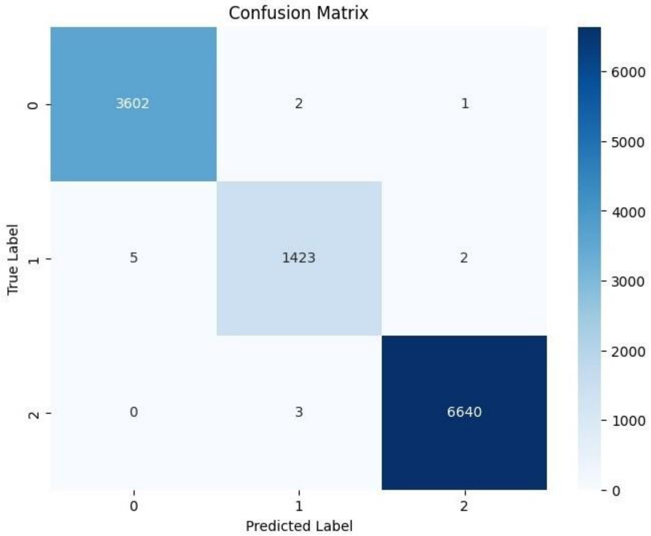

Fig. 5. Confusion Matrix of Tuna using KNN

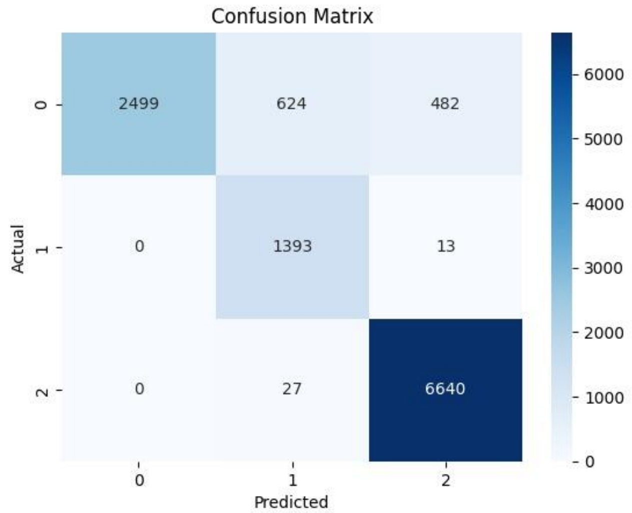

Fig. 6. Confusion Matrix of Tuna using Naive Bayes

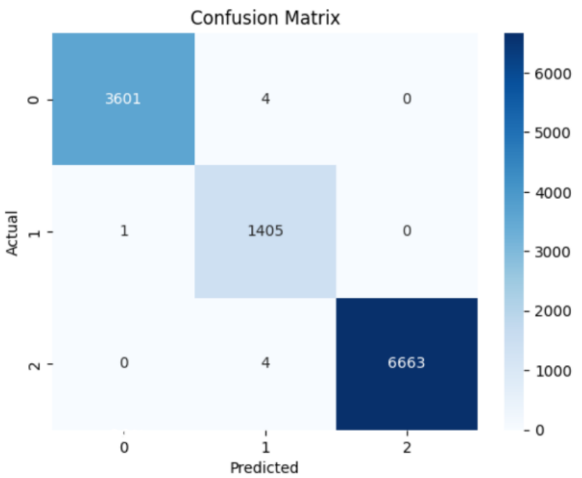

Fig. 7. Confusion Matrix of Tuna using SVM

Table IV. presents the evaluation results on tuna classification, where the KNN and SVM algorithm demonstrates better performance compared to the Naive Bayes algorithm. SVM and KNN achieves an impressive accuracy of 99%, indicating that it accurately classifies almost all instances with minimal error. In contrast, the Naive Bayes algorithm shows a noticeably lower accuracy of 90%, suggesting a higher margin for misclassification.

TABLE IV. CLASSIFICATION REPORT TUNA

Gradient	Class	Precision	Recall	F1-Score	Support
KNN	Fresh	1.00	1.00	1.00	3605
	Less Fresh	1.00	1.00	1.00	1430
	Not Fresh	1.00	1.00	1.00	6643
	Accuracy			1.00	11678
	Macro avg	1.00	1.00	1.00	11678
	Weighted avg	1.00	1.00	1.00	11678
Naive Bayes	Fresh	1.00	0.69	0.82	3605
	Less Fresh	0.68	0.99	0.81	1406
	Not Fresh	0.93	1.00	0.96	6667
	Accuracy			0.90	11678
	Macro avg	0.87	0.89	0.86	11678
	Weighted avg	0.92	0.90	0.90	11678
SVM	Fresh	1.00	1.00	1.00	3605
	Less fresh	0.99	1.00	1.00	1406
	Not fresh	1.00	1.00	1.00	6667
	Accuracy		1.00	11678	
	Macro avg	1.00	1.00	1.00	11678
	Weighted avg	1.00	1.00	1.00	11678

Moreover, when analyzing the performance on the training data, KNN and SVM maintains its high effectiveness, exhibiting perfect precision, recall, and F1 scores, all reaching the ideal value of 1. This reflects ability to consistently and precisely identify and classify each sample of tuna fish, minimizing both false positives and false negatives. On the other hand, the Naive Bayes algorithm yields lower precision, recall, and F1 scores, indicating it is less capable of identifying the correct class in certain instances.

These results highlight the overall reliability of the KNN and SVM algorithm in tuna classification, as it successfully captures the data with minimal classification errors. In comparison, the Naive Bayes algorithm, though still relatively accurate, falls short in its predictive capabilities, resulting in a less optimal performance when compared to KNN and SVM. This demonstrates that for this specific dataset, KNN and SVM is the more suitable algorithm for achieving high classification accuracy and precision.

V. CONCLUSION

This study compares the K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Naive Bayes algorithms to assess and classify tuna freshness based on the gases emitted, using a machine learning model. Gas changes were detected using MQ-135, MQ-9, and MQ-2 sensors. The research evaluated performance using metrics such as accuracy, precision, recall, and F1-score, revealing that the KNN and SVM algorithm achieved a 99% accuracy in classifying tuna freshness, while Naive Bayes demonstrated a lower accuracy of 90%. The KNN and SVM method presents a promising approach to ensuring consumer safety by accurately identifying seafood freshness in real-time. The integration of gas sensors with machine learning algorithms proves highly effective in determining fish freshness and can serve as a foundation for future research. This study contributes to the field of seafood quality assessment by demonstrating the successful combination of gas sensors and machine learning techniques for classifying tuna freshness.

REFERENCES

[1] M. C. H. Soccol and M. Oetterer, "Seafood as functional food," *Brazilian Arch. Biol. Technol.*, vol. 46, no. 3, pp. 443–454, 2003.

[2] Dwitri waluyo, "Ketersediaan Ikan Aman di Ramadan dan Lebaran," INDONESIA.GO. ID Portal Informasi Indonesia. [Online]. Available: https://indonesia.go.id/kategori/editorial/8077/ketersediaan-ikan-aman-di-ramadan-dan-lebaran?lang=1#:~:text=KKP mencatat angka konsumsi ikan,55%2C16 kilogram per kapita

[3] M. Iqbal, & A. N. Rochmah, "Keamanan Pangan: Higiene dan Sanitasi Usaha Jasa Boga," Penerbit Salemba, 2023.

[4] FAO, "Quality and safety of fish and fish products." p. 1, 2015. [Online]. Available: http://www.fao.org/fishery/topic/1514/en

[5] J. D. MacDonald and R. L. Mazany, "Quality Improvement: Panacea for the Atlantic Fishing Industry?," Can. Public Policy / Anal. Polit., vol. 10, no. 3, p. 278, 1984, doi: 10.2307/3550321.

[6] World Trade Organization. (2023). Trade Facilitation and Food Safety: A Case Study of the United States. [Online].

[7] E. Yavuzer, "Determination of fish quality parameters with low cost electronic nose," *Food Biosci.*, vol. 41, no. January, p. 100948, 2021.

[8] D. R. Wijaya, N. F. Syarwan, M. A. Nugraha, D. Ananda, T. Fahrudin, and R. Handayani, "Seafood Quality Detection Using Electronic Nose and Machine Learning Algorithms With Hyperparameter Optimization," *IEEE Access*, vol. 11, no. May, pp. 62484–62495, 2023.

[9] P. Srinivasan, J. Robinson, J. Geevaretnam, and J. B. B. Rayappan, "Development of electronic nose (Shrimp-Nose) for the determination of perishable quality and shelf-life of cultured Pacific white shrimp (Litopenaeus Vannamei)," *Sensors Actuators, B Chem.*, vol. 317, no. April, p. 128192, 2020.

[10] D. Y. Kim, S. W. Park, and H. S. Shin, "Fish Freshness Indicator for Sensing Fish Quality during Storage," *Foods*, vol. 12, no. 9, p. 1801, 2023.

[11] Hanwei Electronics, "Technical MQ-9 Gas Sensor," vol. 1, pp. 3–6, 2018, [Online]. Available: https://www.electronicoscaldas.com/datasheet/MQ-9_Hanwei.pdf

[12] Hanwei Electronics, "Technical Data Mq135 Gas Sensor," Hanwei Electron. Co.,Ltd,

[13] H Hanwei Electronics, "Technical Mq-2 Gas Sensor," *Smoke Sens.*, vol. 1, no. 1, pp. 3–5, 2006.\

[14] J. Tavares et al., "Physical Emerging Technologies," Foods, pp. 1–20, 2021, [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8066737/

[15] van den Berg, R.A., Hoefsloot, H.C., Westerhuis, J.A. *et al.* Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7**, 142 (2006). https://doi.org/10.1186/1471-2164-7-142