

## Evaluación de la precisión de acierto de un conjunto desbalanceado mediante la combinación de clasificadores

Nancy B. Ganz<sup>1</sup>, Alicia E. Ares<sup>1</sup> y Horacio D. Kuna<sup>2</sup>

<sup>1</sup> Instituto de Materiales de Misiones (IMAM-CONICET), Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones, Posadas, Misiones, Argentina.

<sup>2</sup> Instituto de Investigación, Desarrollo e Innovación en Informática (IIDII), Facultad de Ciencias Exactas Químicas y Naturales, Universidad Nacional de Misiones, Posadas, Misiones, Argentina.

{nancy.bea.ganz, a.e.ares, hdkuna}@gmail.com

**Resumen.** El campo de la ciencia de datos ha tenido muchos avances respecto a la aplicación y desarrollo de técnicas en el sector de la salud. Sin embargo, no se han encontrado trabajos que apliquen procesos o técnicas de aprendizaje automático específicamente al campo de la odontología, como es el caso de los injertos o implantes dentales. En este trabajo se estudia la aplicación de varios clasificadores para evaluar la precisión de clasificación de forma integrada sobre un conjunto desbalanceado, referido a historias clínicas de implantes dentales recolectados en puntos característicos de la Provincia de Misiones, Argentina. El modelo abarcó la combinación de los clasificadores: Random Forest, SVC, KNN, Naive Bayes y un Perceptrón Multicapa. La integración de las predicciones se realiza a través del promedio de las mismas, empleando un umbral y pesos en los clasificadores. Como resultado, cada clasificador logró hasta un 68% de acierto de la clase minoritaria, mientras que la integración de todas las predicciones permitió alcanzar el 72% de casos correctamente identificados como fracaso.

**Palabras claves:** ensamble de clasificadores, predicción, clase minoritaria, postoperatorio, implantes dentales.

### 1 Introducción

La combinación de modelos puede ser elemental a la hora de tomar una decisión, debido a que trata de obtener la solución más eficiente para un problema en cuestión. Es posible combinar decisiones obtenidas con el mismo o distintos clasificadores de base [1].

Los métodos de combinación de clasificadores, son aquellos donde dado un conjunto de clasificadores ya entrenados, combinan de distintas formas los resultados para retornar un valor más preciso que el de los clasificadores individuales [2]. La integración suele ser más precisa, porque los datos de entrenamiento pueden no proporcionar información suficiente para elegir el mejor clasificador y en esta situación la combinación es la mejor opción [3].

En este trabajo se estudia la aplicación de varios métodos de clasificación para la predicción del resultado postoperatorio (éxito o fracaso) de un conjunto de datos de implantes dentales. Con el objetivo de aumentar el acierto de la clase minoritaria o de

los fracasos. El conjunto de datos se basa en historias clínicas de pacientes que se han sometido a procesos quirúrgicos de colocación de implantes dentales en la Provincia de Misiones, Argentina. El procedimiento que se propone en este trabajo utiliza los clasificadores: Random Forest (RF) [4], SVC [5], KNN [6], Naive Bayes (NB) [7] y Perceptrón Multicapa (MLP) [8]. El ensamble consistió en aplicar pesos a los clasificadores y promediar sus predicciones. Finalmente, para asignar la etiqueta de clase final se aplicó un umbral.

La contribución de este trabajo es un modelo de aprendizaje automático para la predicción del resultado postoperatorio (éxito o fracaso) en implantes dentales, el cual es un dominio de poco conocimiento. Asimismo, demostramos que los sistemas de múltiples clasificadores también pueden ser aplicados al estudio de caso, ya que permite lograr mejores rendimientos de clasificación que los alcanzados de manera individual por los clasificadores.

El resto del documento está estructurado de la siguiente manera. En la sección 2 se exponen antecedentes en la combinación de clasificadores, en la sección 3 se describe el enfoque propuesto, la sección 4 presenta los resultados experimentales llevados a cabo, y en la sección 5 se exhibe las principales conclusiones y líneas de investigación futuras.

## 2 Trabajos Relacionados

Existen varios trabajos de investigación sobre la combinación o integración de clasificadores, para mejorar el acierto de predicción o inclusive para no sesgar la decisión sobre los resultados de un solo clasificador [9]. A continuación, se presentan y examinan algunos trabajos previos sobre este tema.

Miao et al. [10] proponen un procedimiento para mejorar la precisión en la identificación de genes mediante la integración de los clasificadores SVM, RF y máquina de aprendizaje externo (ELM). Luego del entrenamiento y predicción con los tres clasificadores, los resultados fueron combinados a través del método de votación mayoritaria [9]. Lograron obtener una precisión mayor a través de la integración de las predicciones que de forma individual. De igual manera, Catal y Nengir [11] han utilizado los clasificadores NB y SVM mediante la técnica de voto mayoritario para la integración de las predicciones. Demostraron a través de sus experimentaciones sobre varios conjuntos de datos que los sistemas de clasificadores múltiples mejoran la precisión. Otro trabajo de similares características es el de Pandey y Taruna [12], donde proponen un clasificador integrado utilizando un árbol de decisión J48, KNN y agregación de estimadores de una dependencia (AODE), sobre un conjunto de datos de rendimiento académico de estudiantes de ingeniería. En este modelo, cada clasificador individual genera su valor de predicción y se integran a través del producto de las probabilidades, donde la etiqueta de clase final está representada por el máximo de una probabilidad posterior. Además, Mukhopadhyay et al. [13], utilizan diferentes métodos de combinación para el reconocimiento de imágenes en documentos, con el mismo clasificador de base MLP. Consiguen una precisión máxima del 98,45% con una mejora del 7%, sobre el clasificador individual de mejor desempeño.

Inspirados en las ideas anteriores, proponemos un procedimiento de aprendizaje automático mediante la utilización de múltiples clasificadores para un dominio de poco conocimiento, como es el caso de los implantes dentales. El enfoque propuesto fue capaz de superar el porcentaje de acierto logrado por cada clasificador.

### 3 Materiales y Métodos

La metodología de trabajo que se plantea, consistió en llevar a cabo un conjunto de pasos para obtener el grado de pertenencia del atributo clase. El enfoque propuesto se resume en la Figura 1.

Para el propósito de este trabajo, se utilizó un conjunto de datos de historias clínicas de pacientes que se han sometido a procesos quirúrgicos de colocación de implantes dentales en la Provincia de Misiones, Argentina. Este conjunto de datos estaba representado a través de 4 dimensiones: datos del paciente (antecedentes y condiciones médicas de los pacientes a la hora de la intervención), datos del implante (características del implante utilizado por el especialista implantólogo), datos de la fase quirúrgica (procedimiento de intervención quirúrgica y mejoramiento del lecho óseo del paciente) y datos del seguimiento postoperatorio (resultado del proceso de colocación del implante, es decir si el proceso de oseointegración implante/tejido-óseo tuvo éxito o fracasó). El mismo contaba con un total de 1165 tuplas, 34 características y tenía la particularidad de ser un conjunto desbalanceado, es decir atributo de clase binario con 1009 casos etiquetados como éxito y 156 como fracaso.

Primeramente, se procedió a leer el conjunto de datos y seleccionar la clase objetivo a predecir (característica que representa el resultado del proceso postoperatorio), para pasar a seleccionar las características más relevantes a través de la combinación de los métodos: Information Gain[14], Gain Ratio[15], Random Forest importance[4], Relief[16] y Chi Squared[17]. El procedimiento de combinación de estos métodos se describe en el trabajo [18].

Luego de obtener el subconjunto de características más importantes, se procedió a dividir el conjunto de datos. Una estrategia común consiste en tomar todos los datos etiquetados y dividirlos en subconjunto de entrenamiento y evaluación, normalmente con una proporción del 70 al 80 % para entrenamiento y un 20 al 30 % para evaluación o prueba [19]–[21]. Esta división va a depender en gran medida del número total de muestras y del modelo a entrenar [22], [23]. En nuestro caso se dividió los datos de forma aleatoria para preservar la distribución de ambas clases en: 70 % para entrenamiento y 30 % para evaluación [24]–[27]. Garantizando que todos los casos se encuentren representados en ambos conjuntos.

Conjuntamente, se definió cada uno de los clasificadores seleccionados para el estudio de caso. Todos fueron empleados de forma probabilística. Estos algoritmos de aprendizaje automático poseen hiper parámetros que deben configurarse para lograr su mejor rendimiento. Por esta razón, se precisó:

1. Un espacio de búsqueda: se definió valores para los hiper parámetros y se fue ajustando en función de la medida de rendimiento.

2. Un algoritmo de optimización o ajuste: se empleó el método GridSearchCV [28], es costoso en cuanto a rendimiento, pero permite cubrir todo el espacio de búsqueda.
3. Un método de evaluación: se utilizó validación cruzada de 10 iteraciones.
4. Una medida de rendimiento: se fijó la métrica precisión de equilibrio [29].

Toda la implementación del enfoque propuesto se realizó sobre Python con la librería “Scikit-learn” [30]. En la Tabla 1, se exponen los hiper parámetros que se buscó ajustar para lograr el mejor desempeño de cada uno de los clasificadores, sobre el conjunto de datos utilizado.

**Tabla 1.** Híper parámetros y espacio de búsqueda definido para los clasificadores: RF, SVC, KNN, NB y MLP.

| Clasificadores | Parámetros                | Espacio de búsqueda                    |
|----------------|---------------------------|--|
| RF             | <i>n_estimators</i>       | <i>range (1, 200)</i>                  |
|                | <i>criterion</i>          | <i>gini, entropy</i>                   |
|                | <i>bootstrap</i>          | <i>True, False</i>                     |
| SVC            | <i>kernel</i>             | <i>linear, rbf, poly</i>               |
|                | <i>C</i>                  | <i>range (1, 10)</i>                   |
|                | <i>gamma</i>              | <i>range (1, 10)</i>                   |
|                | <i>degree</i>             | <i>range (1, 10)</i>                   |
| KNN            | <i>n_neighbors</i>        | <i>range (1, 100)</i>                  |
|                | <i>weights</i>            | <i>uniform, distance</i>               |
|                | <i>p</i>                  | <i>manhattan, euclidean</i>            |
| NB             | <i>alpha</i>              | <i>[0, 0.1, 0.2, 0.3, ..., 0.9, 1]</i> |
|                | <i>fit_prior</i>          | <i>True, False</i>                     |
|                | <i>class_prior</i>        | <i>[0.5,0.5], [0.4,0.6], [0.6,0.4]</i> |
| MLP            | <i>hidden_layer_sizes</i> | <i>range (1,5)</i>                     |
|                | <i>activation</i>         | <i>logistic, tanh, relu</i>            |
|                | <i>alpha</i>              | <i>[0.0001, 0.05]</i>                  |
|                | <i>solver</i>             | <i>lbfgs, sgd, adam</i>                |
|                | <i>learning_rate</i>      | <i>constant, invscaling</i>            |

Tanto para el entrenamiento como para la predicción se obtuvieron métricas para evaluar el rendimiento final. Estas métricas fueron: matriz de confusión [31] y exactitud [29].

La integración de las predicciones consistió en multiplicar para cada tupla el valor de probabilidad de cada clase, obtenida por cada clasificador por el peso asignado al mismo. El peso fue determinado mediante una búsqueda en cuadrícula utilizando un parámetro de prueba con valores comprendidos entre 0 y 1. Esta búsqueda fue sometida a una validación cruzada de 10 iteraciones, donde se medió la exactitud [31], [32] de cada clasificador para la clase en cuestión, seleccionando el valor que logró la mejor precisión [33], [34].

Finalmente, se aplicó un umbral para determinar la etiqueta de clase final. Este umbral también estuvo determinado por una búsqueda en cuadrícula, utilizando valores

comprendidos entre 0.1 y 0.5 con incrementos de 0.1 en cada prueba. Se seleccionó el valor del umbral que permitió obtener el mejor resultado de clasificación (matriz de confusión).

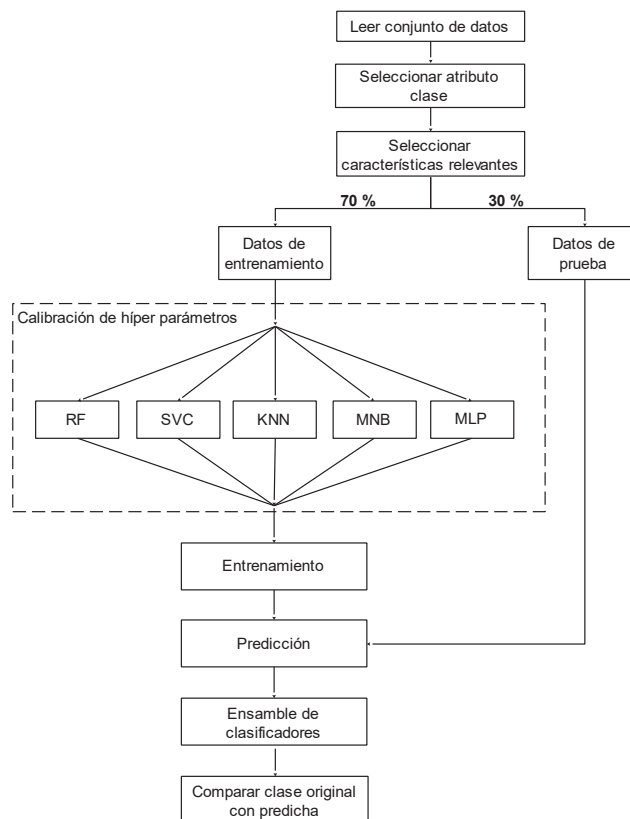
La probabilidad de la etiqueta de clase estuvo dado por:

$$prob[c][i] = ((prob.RF.c * w_1) + (prob.SVC.c * w_2) + (prob.KNN.c * w_3) + (prob.MNB.c * w_4) + (prob.MLP.c * w_5))/5$$

$$prob[predic][i] = np.where(prob[P] \geq \mu, 'FRACASO', 'EXITO')$$

Donde:

- ❖  $i$ : número de tupla.
- ❖  $c$ : etiqueta de clase, E (*EXITO*) o F (*FRACASO*).
- ❖  $prob.[RF | SVC | KNN | NB | MLP].c$ : probabilidad obtenida por el clasificador RF, SVC, KNN, NB o MLP para la clase E o F.
- ❖  $w_1, w_2 \dots w_5$ : peso asignado a cada clasificador en función de su desempeño para  $c$ . Donde  $w \in \mathbb{Q}$ ,  $\{w / 0 \leq w \leq 1\}$ , siendo 0 el peso menor y 1 el mayor.
- ❖  $prob[c][i]$ : valor de probabilidad final de  $c$ , promediada y multiplicada por los pesos definidos para cada clasificador.
- ❖  $\mu$ : valor de probabilidad necesario para considerar a la etiqueta  $c$  como F.
- ❖  $probl[predic][i]$ : etiqueta de clase final asignada en base al umbral  $\mu$ .



**Fig. 1.** Enfoque propuesto para la integración de las predicciones de los clasificadores utilizados sobre el conjunto de datos desbalanceado de implantes dentales.

## 4 Resultados

A continuación, se presentan los resultados logrados en este trabajo de investigación sobre los datos de prueba del conjunto de datos utilizados.

Luego de la selección de características sobre el conjunto de implantes dentales, se logró un subconjunto conformado por 14 características de un total de 34. En la tabla 2 se presentan las matrices de confusión y el porcentaje de acierto (exactitud) obtenido por cada clasificador de forma individual.

**Tabla 2.** Matiz de confusión y exactitud alcanzada por los clasificadores RF, SVC, KNN, NB y MLP sobre el conjunto de datos de implantes dentales.

| Clasificador | EXITO     | FRACASO  | Exactitud |
|--------------|-----------|----------|-----------|
| <i>RF</i>    | 292<br>16 | 8<br>34  | 93%       |
| <i>SVC</i>   | 294<br>17 | 6<br>33  | 93%       |
| <i>KNN</i>   | 294<br>17 | 6<br>33  | 93%       |
| <i>NB</i>    | 250<br>24 | 50<br>26 | 79%       |
| <i>MLP</i>   | 281<br>16 | 19<br>34 | 90%       |

Una vez encontrados los pesos para cada clasificador con la validación cruzada, se multiplicó las predicciones por el peso definido a cada clasificador, se promedió y se aplicó el umbral.

El valor del umbral que permitió obtener el mejor resultado de clasificación fue  $\mu \geq 0.3$ . Es decir, todas aquellas probabilidades mayores o iguales a 0.3 para la clase fracaso fueron consideradas verdaderos fracasos, los casos contrarios fueron considerados como éxitos.

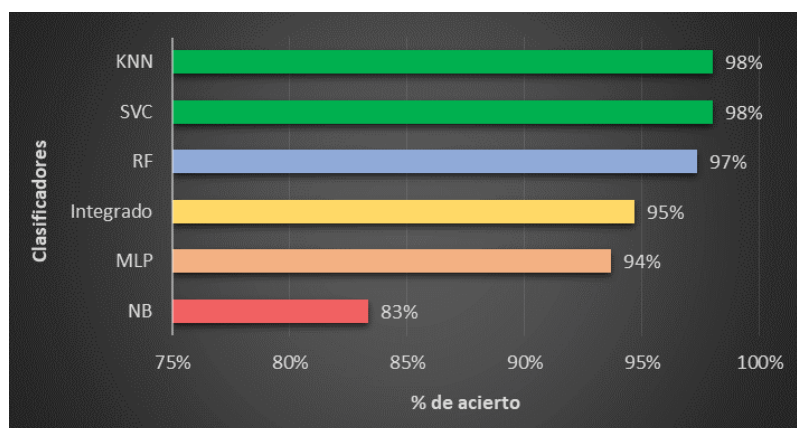
El ensamble permitió confeccionar una matriz de confusión del enfoque propuesto. En la tabla 3 se aprecia que se logró incrementar el acierto de la clase minoritaria (fracaso).

**Tabla 3.** Matiz de confusión y exactitud alcanzada por el enfoque propuesto sobre el conjunto de datos de implantes dentales.

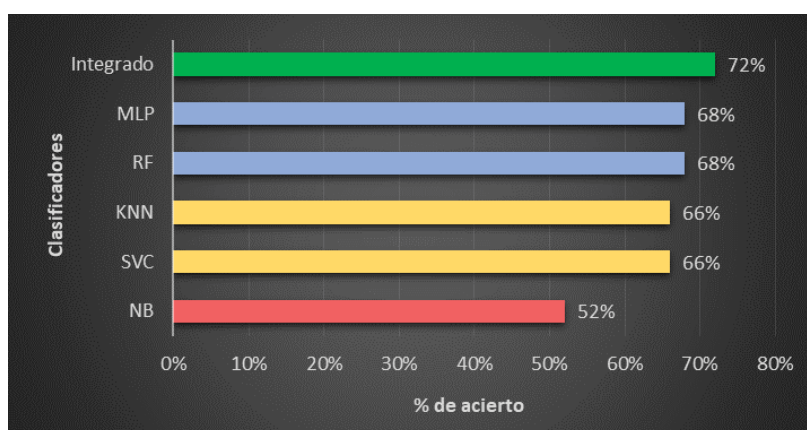
| Clasificador             | EXITO     | FRACASO  | Exactitud |
|--------------------------|-----------|----------|-----------|
| <i>Enfoque propuesto</i> | 284<br>14 | 16<br>36 | 91%       |

En la figura 2, se observa que si bien el rendimiento de la integración de las predicciones no fue la mejor opción para la clase éxito, no quiere decir que éste haya sido el peor en comparación con las predicciones individuales. Mientras que en la figura

3, se aprecia que la integración de las probabilidades fue la mejor opción, ya que permitió obtener el mayor porcentaje de acierto de la clase minoritaria.



**Fig. 2.** Eficiencia de cada clasificador para la clase “EXITO” en comparación con el enfoque propuesto.



**Fig. 3.** Eficiencia de cada clasificador para la clase “FRACASO” en comparación con el enfoque propuesto.

Mediante la experimentación realizada sobre el conjunto de datos de implantes dentales, la integración de las probabilidades de los clasificadores logró el mejor porcentaje de acierto de la clase minoritaria, en comparación con el rendimiento de los clasificadores de forma individual. Donde se logró un acierto del 72% de forma integrada, mientras que de forma individual los clasificadores lograron alcanzar un máximo del 68% de acierto de los fracasos. Si bien la exactitud del enfoque propuesto no fue el mejor, fue el que más fracasos acertó.

Los expertos consultados en patologías bucales y rehabilitación compleja en implantología oral, de distintos puntos de la provincia de Misiones, Argentina,

coincidieron y remarcaron que en este campo de estudio es menos delicado etiquetar un caso como fracaso, que etiquetarlo como éxito cuando era un eventual fracaso.

## 5 Conclusiones y futuras líneas de investigación

Las técnicas de clasificación se basan en conceptos de estimación diferentes, tratando siempre de asociar las mejores características de cada uno de ellos, combinándolos de alguna forma. La finalidad de este trabajo fue la aplicación de varios clasificadores, para predecir el resultado del proceso postoperatorio de un conjunto de implantes dentales y tratar de aumentar el acierto de la clase minoritaria. Se buscó demostrar que en ocasiones es mejor integrar las predicciones de los clasificadores, para no sesgar la decisión sobre un solo resultado. Además, utilizar predicciones integradas permite conocer diversos resultados para un mismo caso, ya que se utilizan más de un clasificador y esto posibilita una asignación de etiqueta o clasificación más precisa.

Como resultado, cada clasificador logró hasta un 68% de acierto para la clase minoritaria, mientras que la integración de todas las predicciones permitió alcanzar el 72% de casos correctamente identificados como fracasos.

Se alcanzó apreciar que si bien KNN y SVC fueron los clasificadores que mejor acierto de la clase éxito lograron, no fue así para la clase fracaso. De la misma manera MLP y RF fueron los clasificadores que mejor acierto de fracasos alcanzaron, no siendo así para la clase éxito. Por este motivo es mejor tener la opinión o predicción de varios modelos, para comparar resultados a la hora de la toma de decisiones.

En cuanto al clasificador NB se valoró que su rendimiento o porcentaje de acierto fue el menor para ambas clases. Se buscó calibrar y optimizar todo lo posible, y aun así no se lograron resultados comparables a los demás clasificadores.

Finalmente, se plantea como trabajo futuro validar el enfoque propuesto con otros conjuntos de datos del área de la salud o la medicina. Conjuntamente, se plantea validar con conjuntos de datos artificiales. Además, se propone la inclusión o ampliación de los clasificadores utilizados, para evaluar la posibilidad de ajustar el porcentaje de acierto de ambas clases.

## Referencias

- [1] Y. Lu, "Knowledge integration in a multiple classifier system," *Appl. Intell.*, vol. 6, no. 2, pp. 75–86, 1996.
- [2] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, 2nd ed. John Wiley & Sons, 2014.
- [3] M. Mohandes, M. Deriche, and S. O. Aliyu, "Classifiers Combination Techniques: A Comprehensive Review," *IEEE Access*, vol. 6, pp. 19626–19639, 2018.
- [4] L. Breiman, "Random Forest," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [5] C. Chang and C. Lin, "LIBSVM: A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–39, 2011.
- [6] N. S. Altman, "An Introduction to Kernel and Nearest-Neighbor



- Nonparametric Regression,” *Am. Stat.*, vol. 46, no. 3, pp. 175–185, 1992.
- [7] C. D. Manning, P. Raghavan, and H. Schütze, “Text classification and Naive Bayes,” in *Introduction to Information Retrieval*, Cambridge University Press, 2009, pp. 253–287.
- [8] B. Irie and Sei Miyake, “Capabilities of Three-layered Perceptrons,” *IEEE International Conf. Neural Networks*, pp. 641–648, 1988.
- [9] J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho, “Multiple classifiers in biometrics. part 1: Fundamentals and review,” *Inf. Fusion*, vol. 44, no. December 2017, pp. 57–64, 2018.
- [10] Y. Miao, H. Jiang, H. Liu, and Y. dong Yao, “An Alzheimers disease related genes identification method based on multiple classifier integration,” *Comput. Methods Programs Biomed.*, vol. 150, pp. 107–115, 2017.
- [11] C. Catal and M. Nangir, “A sentiment classification model based on multiple classifiers,” *Appl. Soft Comput. J.*, vol. 50, pp. 135–141, 2017.
- [12] M. Pandey and S. Taruna, “Towards the integration of multiple classifier pertaining to the Student’s performance prediction,” *Perspect. Sci.*, vol. 8, pp. 364–366, 2016.
- [13] A. Mukhopadhyay, P. Singh, R. Sarkar, and M. Nasipuri, “A Study of Different Classifier Combination Approaches for Handwritten Indic Script Recognition,” *J. Imaging*, vol. 4, no. 2, pp. 1–21, 2018.
- [14] C. E. Shannon, “A Mathematical Theory of Communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
- [15] J. R. Quinlan, “Induction of Decision Trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
- [16] K. Kira and L. A. Rendell, “A Practical Approach to Feature Selection,” *Mach. Learn. Proc.*, pp. 249–256, 1992.
- [17] K. Pearson, “On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling,” *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 50, no. 302, pp. 157–175, 1900.
- [18] N. B. Ganz, F. A. Domínguez, A. E. Ares, and H. D. Kuna, “Selección de características mediante la combinación de métodos para evaluar la precisión de clasificación en un conjunto de datos de implantes dentales,” *XXI Workshop de Investigadores en Ciencias de la Computación*. San Juan, Argentina, pp. 263–267, 2019.
- [19] G. Singh, “Surface Treatment of dental implants: A review,” *IOSR J. Dent. Med. Sci.*, vol. 17, no. 2, pp. 49–53, 2018.
- [20] M. Moran and G. Gordon, “Curious Feature Selection,” *Inf. Sci. (Ny)*, vol. 485, pp. 42–54, 2019.
- [21] I. Sumaiya Thaseen and C. Aswani Kumar, “Intrusion detection model using fusion of chi-square feature selection and multi class SVM,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 4, pp. 462–472, 2017.
- [22] X. Fan and H. Shin, “Road vanishing point detection using weber adaptive local filter and salient-block-wise weighted soft voting,” *IET Comput. Vis.*, vol. 10, no. 6, pp. 503–512, 2016.
- [23] Y. Quan, Y. Xu, Y. Sun, and Y. Huang, “Supervised dictionary learning with multiple classifier integration,” *Pattern Recognit.*, vol. 55, pp. 247–260, 2016.

- [24] B. T. Pham, M. D. Nguyen, K. T. T. Bui, I. Prakash, K. Chapi, and D. T. Bui, “A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil,” *Catena*, vol. 173, no. September 2018, pp. 302–311, 2019.
- [25] G. Isabelle, W. Maharani, and I. Asror, “Analysis on Opinion Mining Using Combining Lexicon-Based Method and Multinomial Naïve Bayes,” *2018 Int. Conf. Ind. Enterp. Syst. Eng. (ICoIESE 2018)*, vol. 2, no. IcoIESE 2018, pp. 214–219, 2019.
- [26] K. Bhattacharjee and M. Pant, “Hybrid Particle Swarm Optimization-Genetic Algorithm trained Multi-Layer Perceptron for Classification of Human Glioma from Molecular Brain Neoplasia Data,” *Cogn. Syst. Res.*, vol. 58, pp. 173–194, 2019.
- [27] D. Chong, N. Zhu, W. Luo, and X. Pan, “Human thermal risk prediction in indoor hyperthermal environments based on random forest,” *Sustain. Cities Soc.*, vol. 49, no. April, p. 101595, 2019.
- [28] scikit-learn, “Tuning the hyper-parameters of an estimator,” 2019. [Online]. Available: [https://scikit-learn.org/stable/modules/grid\\_search.html#grid-search](https://scikit-learn.org/stable/modules/grid_search.html#grid-search). [Accessed: 04-Jul-2019].
- [29] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
- [30] scikit-learn, “scikit-learn: Machine Learning in Python,” 2019. [Online]. Available: <https://scikit-learn.org/stable/>. [Accessed: 04-Jul-2019].
- [31] R. Susmaga, “Confusion Matrix Visualization,” in *Intelligent Information Processing and Web Mining*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 107–116.
- [32] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [33] D. Ruano-Ordás, I. Yevseyeva, V. B. Fernandes, J. R. Méndez, and M. T. M. Emmerich, “Improving the drug discovery process by using multiple classifier systems,” *Expert Syst. Appl.*, vol. 121, pp. 292–303, 2019.
- [34] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-garadi, “Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions,” *Inf. Fusion*, vol. 46, no. June 2018, pp. 147–170, 2019.