ISSN 1870-4069

Estudio del desbalance de clases en bases de datos de microarrays de expresión genética mediante técnicas de Deep Learning

H. Cruz-Reyes[1], A. Reyes-Nava[2], E. Rendón-Lara[1], R. Alejo[1]

[1] Instituto Tecnológico de Toluca,
México

[2] Universidad Autónoma del Estado de México
México

{lugotache,adriananava0}@gmail.com, ralejoll@hotmail.com, erendonl@toluca.tecnm.mx

Resumen. Hoy en día, se ha incrementado significativamente el interés por desarrollar aplicaciones de Deep Learning enfocadas a atender problemas de aprendizaje automático, reconocimiento de patrones y minería de datos en el contexto del Big Data. Esto se debe principalmente a la alta capacidad de procesamiento y buen rendimiento en la selección de características, predicción y tareas de clasificación que presentan los algoritmos de Deep Learning, además, los algoritmos con este enfoque han mostrado tener un buen desempeño tratando con bases de datos utilizadas en el reconocimiento de imágenes y lenguaje natural, las cuales se caracterizan por ser bases de datos de muy alta dimensionalidad y una notable cantidad de datos o muestras. No obstante, al día de hoy uno de los principales retos abarca la clasificación de bases de datos de alta dimensionalidad, con muy pocas muestras y un alto desbalance de clases, un ejemplo de ello son las bases de datos de microarrays de expresión genética. La principal aportación de este trabajo es el estudio de bases de datos de expresión genética, utilizando los métodos convencionales de Sub y Sobre Muestreo para el balance de clases tales como RUS, ROS y SMOTE, las bases de datos fueron modificadas aplicando un incremento en su desbalance y en otro caso generando ruido artificial para realizar un estudio más detallado.

Palabras clave: Deep learning, perceptrón multicapa, microarrays de expresión genética.

Study of the Imbalance of Classes in Gene Expression Microarray Databases using Deep Learning

Abstract. Today, there has been a significant increase in the interest in developing Deep Learning applications focused on identify problems of automatic learning, pattern recognition and data mining in the Big Data context. This is mainly due to the high processing capacity and good performance in the

selec-tion of characteristics, prediction and classification tasks presented by the Deep Learning algorithms, in addition, the algorithms with this approach have shown to have a good performance with databases used in the recognition of images and natural language, which are characterized as very high dimensionality data-bases and a considerable amount of data or samples. However, today one of the main challenges involves the classification of high dimensionality databases, with very few samples and a high level of class imbalance, an example of which are the microarray databases of genetic expression. The main contribution of this work is the study of databases of genetic expression, using the conventional Over and Under Sampling methods for the balance of classes such as RUS, ROS and SMOTE, in addition the databases were modified applying an increase in its imbalance and generating artificial noise to perform a detailed study.

Keywords: Deep learning, multilayer perceptron, gene expression microarrays.

1. Introducción

Recientemente se ha popularizado el uso de Redes Neuronales Artificiales (ANN por sus siglas en inglés) para llevar a cabo tareas de clasificación enfocadas a problemas reales, una de las redes más usada es el Perceptrón Multicapa (MLP por sus siglas en inglés) entrenado con el método Back-Propagation [1,2], es una de las redes más populares debido a las ventajas que presenta, como son: rapidez, paralelismo inherente, no requiere conocimiento a priori de la distribución estadística de los datos y la tolerancia a fallos.

Tradicionalmente las ANN están conformadas de tres capas (capa de entrada, capa oculta y capa de salida), sin embargo, hoy en día cuando una ANN está compuesta por más de tres capas es conocida como una ANN Deep Learning (DL) [3]. La arquitectura de ANN-DL más representativa es el MLP con varias capas ocultas [3-5]. Las principales ventajas de este tipo de arquitectura son tres: alto desempeño, robustez al sobre-entrenamiento y alta capacidad de procesamiento.

Hoy en día los algoritmos de DL muestran buenos resultados al ser utilizados en la solución de distintos problemas [6-11], los cuales tienen características similares como la gran cantidad de datos y alta dimensionalidad. Sin embargo, uno de los principales retos que actualmente surge es la clasificación de bases de datos de alta dimensionalidad, con muy pocas muestras y alto desbalance de clases. Las bases de datos biomedicas de microarrays de expresión genética tienen las características antes mencionadas, a menudo presentan problemas de desbalance de clases, tienen pocas muestras y alta dimensionalidad. El problema del desbalance de clases surge cuando el conjunto de muestras perteneciente a una clase es mucho mayor al conjunto de muestras de la otra u otras clases [12], este problema se ha identificado como uno de los principales retos de los algoritmos aplicados en el contexto del Big Data [7,10, 11, 13].

Los clasificadores convencionales como el MLP fueron diseñados para trabajar con bases de datos balanceadas, es decir, que el número de muestras sea la misma para cada clase. Por esta razón, cuando se trabaja con una base de datos desbalanceada no se logran resultados óptimos, debido al alto porcentaje de muestras mal clasificadas en las clases menos representadas o minoritarias [14].

Aunado a ello, el método Back-Propagation (utilizado para entrenar el MLP) también se ve afectado por el desbalance, ya que ralentiza la convergencia de la red [15], lo que potencializa una de las desventajas de este método de entrenamiento.

El desbalance de clases ha sido ampliamente estudiando en problemas de dos clases [16], sin embargo, los problemas de múltiples clases [17, 18] y DL han sido poco abordados. Los trabajos enfocados al desbalance en múltiples clases comúnmente utilizan costos asociados a las diferentes clases en la etapa de entrenamiento, a pesar de ello, este enfoque solo es adecuado en el entrenamiento con Back-Propagation cuando el entrenamiento es por lotes o "batch mode" [19]. En la resolución de problemas reales el entrenamiento por lotes es menos usado que el entrenamiento estocástico, ya que este último es usualmente más rápido, alcanza mejores soluciones y puede ser usado para identificar cambios al pasar las muestras por el MLP.

Tradicionalmente los métodos utilizados para atacar el desbalance de clases se basan en duplicar o eliminar muestras hasta alcanzar un equilibrio en el número de muestras por clase, por ejemplo, Random Over-Sampling (ROS) y Random Under-Sampling (RUS) [20]. Uno de los métodos comúnmente usados es Synthetic Minority Over-sampling Technique (SMOTE), propuesto por Chawla et al. [21], esta técnica genera nuevas muestras sintéticas interpoladas en las muestras de la clase minoritaria. Este método ha servido como base de otros métodos de muestreo como Bordeline-SMOTE, Adaptative Synthetic Samplig (ADASYN), SMOTE editing nearest neighbor, entre otros [14, 16].

Por otro lado en las técnicas de sub-muestreo (under-sampling), RUS ha sido reportado como una de las técnicas más efectivas [22]. En estas técnicas se han propuesto métodos caracterizados por incluir un mecanismo heurístico en su funcionamiento, el cual tiene como objetivo eliminar o cambiar las etiquetas de las muestras, ya sean ruido, atípicos o redundantes [23], como ejemplo se puede mencionar a los métodos Neighborhood Clearing Rules y One-sided Selection.

Actualmente ha surgido el interés por desarrollar métodos de muestreo dinámico en el contexto de los MLP, donde el objetivo es usar el número apropiado de muestras o aumentar el tamaño de la clase minoritaria al momento de entrenar el perceptrón. Como ejemplo está el método SNOBALL [24], donde la clase mayoritaria se incrementa gradualmente, otro ejemplo es DyS [25], el cual atenúa el desbalance por medio de un mecanismo de sobre-muestreo e identificación de las muestras difíciles de aprender, es decir, pone más atención a las muestras más difíciles de clasificar.

En general el desbalance de clases afecta negativamente al desempeño de los algoritmos de aprendizaje automático (ML por sus siglas en inglés). El desbalance también está presente el contexto el Big Data donde el uso de Maquinas de Boltzman, redes de creencia, redes convolucionales y en general redes con un enfoque DL han mostrado buenos resultados [10,11], mientras que los algoritmos de aprendizaje automático han mostrado notables deficiencias en su desempeño.

En el contexto de Big Data son pocas las propuestas para atacar el problema del desbalance, actualmente se están adaptando los algoritmos de aprendizaje automático para desempeñarse en este enfoque. El objetivo de este trabajo es identificar el comportamiento del clasificador utilizando bases de datos de microarrays de expresión genética las cuales cuentan con pocas muestras, alta dimensionalidad y en algunos casos desbalance de clases.

2. Trabajos relacionados

Hoy en día en áreas como ML, el reconocimiento de patrones (PR por sus siglas en inglés), y la minería de datos (DM por sus siglas en inglés) se ha demostrado que el desbalance de clases es un problema crucial que afecta la eficiencia del algoritmo [21, 26], incluso en Big Data se considera como uno de los principales retos [11, 12, 14,15]. El desbalance se presenta cuando una o más clases tienen un menor número de muestras (clase minoritaria) que las demás (clase o clases mayoritarias) y afecta directamente a la capacidad de generalización del clasificador ya que este asume que los conjuntos de muestras por clase están balanceados.

El desbalance de clases surge cuando una o más clases se encuentran menos representadas en su número de muestras, en comparación con el número de muestras de otras clases. En las redes neuronales artificiales, en particular, en el MLP, el desbalance acentúa las debilidades de este clasificador, sobre todo cuando el entrenamiento se hace mediante el algoritmo del Back-Propagation [15].

Usualmente las bases de datos pueden estar agrupadas en dos clases o más, el dominio de dos clases ha sido ampliamente estudiado, sin embargo, trabajar con múltiples clases sigue siendo un reto del aprendizaje automático, la minería de datos y el reconocimiento de patrones [17, 18]. Actualmente, el problema del desbalance de clases se ha abordado de muchas maneras y enfoques diferentes, los más estudiados han sido los métodos de muestreo enfocados al desbalance entre clases, estos métodos suelen ser eficaces y son independientes del clasificador [22].

Los métodos de muestreo pueden ser simples y claros como el sobre o sub muestreo aleatorio (ROS o RUS) [14]. El primero replica aleatoriamente muestras existentes en la clase minoritaria, lo que en algunos casos podría dar pie a que ocurra un sobre ajuste [21], y el segundo quita un determinado número de muestras permitiendo así un balance entre el número de elementos por clases, aunque, en algunos escenarios, sería inapropiado debido a la enorme pérdida de información en la base de datos. Por lo tanto, se han desarrollado otros métodos de muestreo "inteligentes" que incluyen un mecanismo heurístico, como SMOTE, el cual crea muestras artificiales de la clase minoritaria mediante la interpolación de muestras existentes cerca de ellas [21] y de esta forma evitar la sobre especialización.

Una técnica propuesta para optimizar las deficiencias de las técnicas de muestreo como ROS o SMOTE es Borderline-SMOTE [27], la cual selecciona muestras de la clase minoritaria que están en el límite, realizando sólo SMOTE en esas muestras. Por otro lado el muestreo sintético adaptativo (ADASYN) es también una extensión de SMOTE, el cual crea en el límite de la región más muestras entre las dos clases que en el interior de la clase minoritaria. SMOTE Editing Nearest Neighbor (ENN) consiste en aplicar SMOTE y, a continuación, la regla ENN. Safe-Level-SMOTE genera muestras sintéticas de la clase minoritaria situadas más cerca del mayor nivel de seguridad, entonces todas las muestras sintéticas sólo se generan en regiones seguras [28]. SMOTE + Tomek Links (TL) es la combinación de SMOTE y TL [20], Neighborhood Cleaning Rule usa la regla ENN, pero sólo elimina las muestras de la clase mayoritaria. Condensed Nearest Neighbor rule (CNN) y One-sided Selection eliminan las muestras redundantes, pero esta última usa TL.

Se han presentado métodos más sofisticados para tratar el problema de desbalance de múltiples clases. Uno de ellos es el costo sensitivo (CS), el cual, es de los temas más relevantes en la investigación del aprendizaje de automático y es una buena solución para el problema de desbalance de clases [13]. El CS utiliza los costos asociados con la clasificación errónea de las muestras, emplea varias matrices de costos que definen los costos de clasificación errónea de cualquier muestra de datos [21]. Sin embargo, en estos métodos, el costo de clasificación errónea debe ser conocido de antemano, pero en un problema de clasificación real, el costo de clasificación errónea es a menudo desconocido. Zhi-Hua y Xu-Ying [29] proporcionan un marco unificado para el uso de CS para abordar el desbalance de clases.

Recientemente, se han propuesto métodos de muestreo dinámico para resolver el problema del desbalance de múltiples clases, los cuales establecen automáticamente la tasa de muestreo, por ejemplo, Fernández-Navarro et al. [30] combinan métodos a nivel de datos utilizando algoritmos genéticos para obtener la mejor relación de sobremuestreo. Alejo et al. [31] usa el error cuadrático medio con este propósito. Chawla et al. [32] proponen un paradigma Wrapper que descubre automáticamente la cantidad de sub-muestreo y tasa de sobre-muestreo para un conjunto de datos basado en la optimización de las funciones de evaluación. En [25] se propone un nuevo algoritmo para determinar el nivel de equilibrio de clases, y además incluyen un mecanismo de selección de patrones de entrenamiento difíciles de aprender, con el propósito de mejorar la capacidad de generalización del MLP entrenado con el algoritmo Back-Propagation. Un método más antiguo de muestreo dinámico (SNOWBALL) es propuesto por Wang y Jean [24] para entrenar redes del tipo MLP con datos desbalanceados, básicamente este método repite el entrenamiento de las muestras de las clases minoritarias hasta que el clasificador las identifica adecuadamente. Por otra parte, trabajos recientes muestran interés en encontrar las mejores muestras para construir el clasificador, por ejemplo eliminando las muestras que son difíciles de aprender o cercanas a la frontera de decisión, ya que podrían ser muestras con "ruido" o "solapadas".

El desbalance de clases es un problema muy frecuente en tareas de visión por computadora, por ejemplo, en el dominio de las redes convolucionales se han usado técnicas de re-muestreo o aprendizaje sensible a costos para hacer frente a este problema.

3. Microarrays de expresión genética

Los microarrays de expresión genética son conjuntos de datos de perfiles de expresión del mundo real que se utilizan en varios tipos de investigación del cáncer. Las bases de datos de este estudio serán Prostate, Ovarian y Breast, las cuales tienen muy pocas muestras, son de alta dimensionalidad y están agrupadas en dos clases, cabe mencionar que el número de elementos por clase no es balanceado. Estas se pueden obtener del Repositorio del conjunto de datos biomédicos de Kent Ridge (http://leo.ugr.es/elvira/DBCRepository/).

Las bases de datos de microarrays de expresión genética generalmente cuentan con un limitado número muestras, tienen una alta dimensionalidad y en algunos casos presentan desbalance de clases.

4. Perceptrón Multicapa Deep Learning (MLP DL)

Un MLP consiste en una red que contiene una capa de entrada, una de salida y una o más capas ocultas [5]. En la capa de entrada se ingresan los datos que serán analizados, estos pasan a la primera capa oculta, los resultados de esta capa pasan a la segunda capa oculta (si es el caso), y al final pasan a la capa de salida; al paso de cada una de las capas ocultas se van asignando pesos sinápticos. El número de nodos de cada capa oculta puede variar [3, 4].

La principal diferencia entre las arquitecturas de aprendizaje automático y aprendizaje profundo es el número de capas ocultas, convencionalmente en el aprendizaje automático las arquitecturas están compuestas de una capa de entrada, una oculta y una de salida, a diferencia del aprendizaje profundo donde se componen de más de 3 capas, debido a que tienen más de una capa oculta. Cuando la red tiene más de 3 capas en su arquitectura se clasifica como aprendizaje profundo [3].

Muchos de los avances del aprendizaje profundo dependen en gran medida de la tecnología que se usa para implementarse, algunas de las librerías más usadas para este enfoque son Theano, PyLearn2, Caffe, Tensorflow y la plataforma de trabajo Apache Spark.

En este trabajo se utilizó una red neuronal de enfoque DL, con una configuración muy similar para todas las bases de datos, donde la única diferencia es la capa de entrada debido a que las bases de datos no tienen el mismo número de atributos. Para ejecutar la red se buscó un entorno que pudiera realizar el procesamiento en paralelo, por lo que se utilizó la plataforma Spark, ya que ofrece un buen desempeño al procesar datos de gran tamaño y alta dimensionalidad, optimizando el tiempo de ejecución y generando resultados confiables.

5. Diseño experimental

El propósito de este trabajo es estudiar el comportamiento del clasificador MLP DL utilizando bases de datos de alta dimensionalidad, pocos patrones y alto desbalance de clases, como lo son las bases de datos de microarrays de expresión genética. Para este trabajo, se utilizaron bases de datos públicas sobre datos de cáncer disponibles en el repositorio Kent Ridge Biomedical Data Set. Los detalles de las bases de datos pueden ser consultados en la Tabla 1.

En la Tabla 1 se puede observar que además de las bases mencionadas se describen dos bases de datos más por cada una, las cuales se pueden identificar porque el nombre de la base de datos incluye los términos “–Ruido” y “–Disminuido”.

Tabla 1. Descripción de las bases de datos.

Base de Datos	Características	No. de Ejemplos	Clase 1	Clase 0
Ovarian	15154	253	162	91
Ovarian-Ruido	15154	253	172	81
Ovarian-Disminuido	15154	243	162	81
Prostate	12600	136	77	59
Prostate -Ruido	12600	136	87	49
Prostate -Disminuido	12600	126	77	49
Breast	24481	97	46	51
Breast -Ruido	24481	97	36	61
Breast -Disminuido	24481	87	36	51

Estas bases de datos se generaron para obtener un desbalance de clases aún más significativo. Para las bases identificadas como "Disminuido" se sustrajeron aleatoriamente 10 muestras, logrando que la clase minoritaria disminuyera su tamaño y así tener un desbalance más relevante. En las bases identificadas como "Ruido" se buscó la manera de generar Ruido Artificial, lo cual se logró seleccionando aleatoriamente diez muestras de la clase minoritaria para cambiar su clase, de este modo disminuir la clase minoritaria y aumentar la clase mayoritaria.

La configuración utilizada en la red neuronal es la establecida por defecto en el MLP de Spark, no obstante, se ajustaron algunos parámetros tales como la capa de entrada con 12600 nodos para la base de datos Prostate, 15154 para Ovarian y 24481 para Breast. Dos capas ocultas (la primera con 90 nodos, la segunda con 80) y finalmente una capa de salida de dos nodos. La configuración usada en las capas ocultas y la capa de salida fue exactamente la misma para todas las bases de datos.

Se aplicó el método Hold-Out [33] para la segmentación de los conjuntos de entrenamiento y test, quedando 60 y 40 por ciento respectivamente, se repitió el proceso de división 10 veces de forma aleatoria, donde, cada conjunto contenía diferentes muestras, es decir, las muestras contenidas en el conjunto de entrenamiento no se encontraban en el de test y viceversa.

Para evaluar la eficacia del modelo se usó el área bajo la curva (AUC por sus siglas en inglés), la cual es una medida ampliamente utilizada en investigaciones sobre desbalance de clases. Para la ejecución se optó por procesar diez veces cada una de las bases de datos y finalmente obtener el promedio de la métrica.

6. Resultados

La Tabla 2 muestra los resultados obtenidos por el MLP DL utilizando AUC para medir la eficacia del clasificador. Los resultados se muestran por cada una de las técnicas de muestreo usadas en todas las bases de datos, los valores en negritas señalan la técnica de muestreo (sub o sobre muestreo) con mejor desempeño en cada una de las bases de datos.

Tabla 2. Resultados obtenidos por el MLP DL utilizando la métrica AUC.

Base de Datos	ORIG	ROS	RUS	SMOTE
Ovarian	0.9634	**0.9792**	0.9624	0.9782
Ovarian-Ruido	0.8859	**0.9418**	0.8930	0.9361
Ovarian-Disminuido	0.9819	**0.9902**	0.9488	0.9896
Prostate	0.8347	0.8614	0.8380	**0.8956**
Prostate -Ruido	0.7121	0.8226	0.7743	**0.8264**
Prostate – Disminuido	0.8533	**0.8971**	0.7982	0.8783
Breast	0.5835	0.6609	0.5612	**0.6690**
Breast -Ruido	0.5851	0.6691	0.6704	**0.6980**
Breast -Disminuido	0.5791	**0.7124**	0.5801	0.6996

Para garantizar resultados fiables, se presenta el puntaje promedio de 10 ejecuciones en cada uno de los experimentos, se utilizaron cifras con cuatro decimales para dar una mayor información de los resultados.

Las bases de datos Ovarian, Prostate y Breast son las bases de datos originales, y en general tienen una buena clasificación, sin embargo, "Breast" presenta una exactitud de 0.5835 que es un nivel de eficacia muy bajo, cabe señalar que es la base con el menor número de muestras y mayor número de atributos, además de presentar el desbalance más pequeño con una diferencia de cinco muestras entre sus clases.

Al aplicar técnicas de muestreo a las bases antes mencionadas, se obtuvieron resultados similares a los de las bases de datos originales cuando se utilizó RUS, por otro lado, al aplicar las técnicas de sobre muestreo ROS y SMOTE se observó una notable mejora en la eficiencia del clasificador. Siendo ROS el que obtuvo la eficiencia más alta en Ovarian, mientras que SMOTE fue el mejor en Prostate y Breast, a pesar de ello, no hubo una diferencia significativa en la eficiencia del clasificador al utilizar ROS y SMOTE.

En las bases de datos Ovarian-Disminuido, Prostate-Disminuido y Breast-Disminuido, donde se incrementó el desbalance de clases sustrayendo diez muestras al conjunto de la clase minoritaria, se observó que elevaron su nivel de eficacia en comparación con las bases de datos originales, RUS volvió a ser el que tiene el más bajo nivel, incluso por debajo de las bases de datos originales. Cabe señalar que la base de datos Breast-Disminuido tuvo un nivel de eficacia considerablemente mayor que el de la base Breast.

En el caso de Ovarian-Ruido, Prostate-Ruido y Breast-Ruido, que son las bases de datos a las que se les aplicó ruido artificial, se observó que todos los métodos de muestreo alcanzaron mejor nivel que la base de datos original. Los resultados de RUS son muy cercanos a los de la base original, por otro lado SMOTE obtuvo la eficiencia más alta en Prostate-Ruido y Breast-Ruido mientras que ROS en Ovarian-Ruido, a pesar de ello, no hubo una diferencia significante en la eficiencia del clasificador al utilizar ROS y SMOTE.

Se puede observar que utilizando técnicas de sobre-muestreo, se obtienen mejores resultados en comparación a las bases de datos originales. En el caso del sub-muestreo aleatorio (RUS) se observa que el desempeño del clasificador es muy similar al de la base de datos original incluso, en algunos casos ligeramente menor, por lo tanto utilizando técnicas tradicionales de sobre muestreo tales como ROS y SMOTE el clasificador podrá obtener mejores resultados al trabajar con bases de datos de alta dimensionalidad, de pocas muestras y alto desbalance de clases.

7. Conclusiones

El desbalance de clases ha sido reconocido como uno de los principales retos a la hora de entrenar clasificadores supervisados, esto debido la mayoría de ellos fueron diseñados para trabajo con bases de datos relativamente balanceadas. Actualmente, muchas de las bases de datos que se están generando presentan problemas de des balance de clases, por ejemplo, las bases de datos de microarray de expresión genética, aunado a esto, características como alta dimensionalidad y escasas de muestras o patrones de entrenamiento caracterizan a estas bases de datos.

El deep learning, ha sido una excelente alternativa para tratar con bases de datos de gran tamaño y dimensionalidad, no obstante, ha mostrado notables deficiencias al trabajar con bases de datos desbalanceadas. En este trabajo se estudió la efectividad de métodos tradicionales para tratar el de balance de clases, en bases de datos de microarrays de expresión genética las cuales se caracterizan por tener pocas muestras o patrones de entrenamiento y un número excesivo de atributos o características.

Resultados presentados este trabajo muestran la efectividad de las técnicas de muestreo ROS y SMOTE para tratar el desbalace de clases en la clasificación de bases de datos de microarrays de expresión genética, sin embargo, se observa una tendencia en SMOTE a producir mejores resultados. Por otro lado, se muestra que es prohibitivo eliminar muestras en este tipo de bases de datos con la finalidad de tratar el desbalance de clases.

Es indudable se requiere profundizar en el tema no solo por la importancia del mismo sino por su relación con otras áreas del conocimiento como la biomedicina y el Big Data. Para trabajos futuros se tiene contemplado estudiar otros algoritmos clásicos para el tratamiento del desbalance de clases y en su momento proponer un nuevo método que ayude a superar las deficiencias de los métodos existentes en el estado del arte.

Referencias

1. Linderman, M., Liu, J., Qi, J., An, L., Ouyang, Z., Yang, J., Tan, Y.: Using artificial neural networks to map the spatial distribution of understorey bamboo from remote sensing data. International Journal of Remote Sensing, 25(9), pp. 1685–1700 (2004)
2. Pal, M.: Extreme learning machine for land cover classification. International Journal of Remote Sensing, 30(14), pp. 3835–3841 (2008)
3. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)

4. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Networks, 61, pp. 85–117 (2015)
5. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature, 521, pp. 436—444 (2015)
6. Nene, S.: Deep learning for natural languaje processing. International Research Journal of Engineering Technology, 4, pp. 930–933 (2017)
7. Inur, A., Ritahani, A., Ahmad, A.: Convolutional Neural Networks and Deep Belief Networks for Analysing Imbalanced Class Issue in Handwritten Dataset. International Journal on Advanced Science Engineering Information Technology, 7, pp. 2302–2307 (2017)
8. Yan, Y., Chen, M., Shyu, M.L., Chen, S.S.: Deep Learning for Imbalanced Multimedia Data Classification. IEEE International Symposium on Multimedia, pp. 483–488 (2015)
9. Mahsereci, E., Ibrikci, T.: Discriminative deep belief networks for microarray based cancer classification. Biomedical Research, 28(3), pp. 1016–1024 (2017)
10. Heureux, A., Grolinger, K., Elyamany, H., Capretz, M.: Machine Learning With Big Data: Challenges and Approaches, In: IEEE Access, 5, pp. 7776–7797 (2017)
11. Zhou, L., Pan, S., Wang, J., Vasilakos, A.V.: Machine learning on big data: Opportunities and challenges. Neurocomputing, 237, pp. 350–361 (2017)
12. Ou, G., Murphey, Y.L.: Multi-class pattern classification using neural networks. Pattern Recognition, 40(1), pp. 4–18 (2007)
13. Salman, H. K.: Cost-Sensitive Learning of Deep Feature Representations From Imbalanced Data. IEEE Transactions on Neural Networks and Learning Systems (2017)
14. López, V., Fernández, A., García, S., Palade, V., Herrera, F.: An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Inf. Sci., 250, pp. 113–141 (2013)
15. Anand, R., Mehrotra, K., Mohan, C., Ranka, S.: An improved algorithm for neural network classification of imbalanced training sets. IEEE Trans Neural Netw, 4, pp. 962–969 (1993)
16. He, H., Garcia, E.: Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), pp. 1263–1284 (2009)
17. Wang, S., Yao, X.: Multiclass imbalance problems: Analysis and potential solutions. IEEE Transactions on Systems, Man, and Cybernetics, 42(4), pp. 1119–1130 (2012)
18. Fernández, A., López, V., Galar, M., De Jesus, M.J., Herrera, F.: Analyzing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. Knowledge-Based Systems, 42(8), pp. 97–110 (2013)
19. Kretzschmar, R., Karayiannis, N.B., Eggimann, F.: Feedforward neural network models for handling class overlap and class imbalance. Int. J. Neural Syst., 15(5), pp. 323–338 (2005)
20. Nguyen, A.B., Phung, S.L.: A supervised learning approach for imbalanced data sets. In: Proc. of the 19th International Conference on Pattern Recognition, pp. 1–4 (2008)
21. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, pp. 321–357 (2002)
22. Japkowicz, N., Stephen, S.: The class imbalance problem: A systematic study. Intell. Data Anal, 6(5), pp. 429–449 (2002)
23. Wilson, D.: Asymptotic properties of nearest neighbor rules using edited data. IEEE Trans. Syst. Man Cyber., 2, pp. 408–420 (1972)
24. Wang, J., Jean, J.: Resolving multifont character confusion with neural networks. Pattern Recognit., 26, pp. 175–187 (1993)
25. Lin, M., Tang, K., Yao, X.: Dynamic sampling approach to training neural networks for multiclass imbalance classification. IEEE Trans. Neural Netw. Learning Syst., 24(4), pp. 647–660 (2013)

26. Debowski, B., Areibi, S., Gréwal, G., Tempelman, J.: A dynamic sampling framework for multi-class imbalanced data. In: ICMLA, (2), pp. 113–118 (2012)
27. Han, H., Wang, W.Y., Mao, B.H.: Borderline-smote: A new over-sampling method in imbalanced data sets learning. In: Proceedings of the 2005 International Conference on Advances in Intelligent Computing, I. (ICIC'05), pp. 878–887 (2005)
28. Bunkhumpornpat, C., Sinapiromsaran, K., Lursinsap, C.: Safe-level-SMOTE: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. In: Proceedings of the 13th Pacific-Asia Conference (PAKDD 2009), 5476, pp. 475–482 (2009)
29. Zhou, Z.H., Liu, X.Y.: On multi-class cost-sensitive learning. Computational Intelligence, 26(3), pp. 232–257 (2010)
30. Fernández-Navarro, F., Hervás-Martínez, C., Antonio Gutiérrez, P.: A dynamic over-sampling procedure based on sensitivity for multi-class problems. Pattern Recogn., 44(8), pp. 1821–1833 (2011)
31. Alejo, R., García, V., Pacheco-Sánchez, J.H.: An efficient over-sampling approach based on mean square error back-propagation for dealing with the multi-class imbalance problem. Neural Processing Letters, 42(3), pp. 603–617 (2014)
32. Chawla, N.V., Cieslak, D.A., Hall, L.O., Joshi, A.: Automatically countering imbalance and its empirical relationship to cost. Data Min. Knowl. Discov., 17, pp. 225–252 (2008)
33. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification. Wiley-Interscience Publication (2001)