

Detección de Niveles de Alzheimer según características extraídas de imágenes médicas y combinación de dataframe de OASIS

Javier Urbina ¹, Diego Segura ¹

¹FACI, UPCH.

Date 2024-05-08

ABSTRACT

La Enfermedad de Alzheimer (EA) es una preocupación mundial que afecta a millones de personas, sin un tratamiento efectivo conocido hasta la fecha. Su detección temprana sigue siendo crucial para gestionar la carga de la AD. Este artículo tiene como objetivo desarrollar un modelo de inteligencia artificial para la detección temprana de la EA mediante el uso de características extraídas de imágenes de resonancia magnética cerebral. El conjunto de datos consiste en imágenes de resonancia magnética cerebral (MRI) utilizadas para evaluar y validar el modelo propuesto de la base de datos OASIS. El método incluye etapas para el preprocesamiento, el entrenamiento del modelo y la validación del modelo, la cual implica evaluar métricas de rendimiento como la precisión, la exactitud y la especificidad, seguido de la optimización de hiperparámetros para mejorar la robustez del modelo. Los resultados experimentales muestran que (será llenado con los resultados finales).

Keywords: Alzheimer's Disease, MRI, Machine Learning, OASIS Database.

envejecimiento normal [5]. Este solapamiento de síntomas dificulta la diferenciación precisa entre estas condiciones, complicando así la intervención temprana y adecuada.

Las imágenes de resonancia magnética (RM) juegan un papel fundamental en la detección de cambios estructurales en el cerebro relacionados con la EA, ofreciendo una ventana para el diagnóstico temprano y la monitorización de la progresión de la enfermedad. La detección temprana de la EA a través de imágenes de RM del cerebro es un reto significativo y crucial para mejorar la gestión y el tratamiento de esta enfermedad neurodegenerativa.

El uso de técnicas avanzadas de aprendizaje automático, aprendizaje profundo y modelos de redes neuronales convolucionales (CNNs) ha demostrado ser prometedor en la clasificación y detección precisa de la enfermedad. Estas tecnologías emergentes tienen el potencial de superar las limitaciones de los métodos tradicionales de diagnóstico, proporcionando herramientas más precisas y eficaces para la detección temprana y el seguimiento de la EA.

1 INTRODUCCIÓN

Aproximadamente 50 millones de personas en todo el mundo padecen demencia [1]. Adicionalmente, en los países de ingresos bajos y medios, el número de afectados está aumentando rápidamente debido al incremento en la esperanza de vida. La forma más común de demencia es la enfermedad de Alzheimer (EA), un trastorno cerebral progresivo que deteriora la memoria y las habilidades cognitivas. Asimismo, se proyecta que el número anual de nuevos casos de Alzheimer y otras demencias se triplicará para el año 2050, alcanzando los 152 millones de casos, lo que equivale a un nuevo caso de demencia cada 3 segundos [1].

Los síntomas de la EA incluyen pérdida severa de la memoria, dificultades con el pensamiento y el razonamiento, problemas para hacer valoraciones y tomar decisiones, así como cambios en la personalidad y el comportamiento [2]. Estos cambios son resultado de la acumulación de placas amiloides y ovillos neurofibrilares en el cerebro, lo que lleva a la atrofia cerebral. A medida que la enfermedad avanza, se observa un daño inicial en el hipocampo y la corteza entorrinal, áreas esenciales para la formación de los recuerdos [3].

Actualmente, no existe una cura para la EA, pero el diagnóstico temprano es crucial para el manejo de los síntomas y la planificación del cuidado [4]. El diagnóstico de la EA se complica considerablemente debido a sus síntomas superpuestos con el

2 ESTADO DEL ARTE (RELATED WORK)

Estudios recientes han implementado modelos como ResNet50 para la extracción automática de características a partir de imágenes de RM, logrando precisiones notables en la clasificación de la enfermedad del Alzheimer, con rangos de precisión que oscilan entre el 85.7%, y el 99% en conjuntos de datos como el MRI ADNI [6]. Además, se ha propuesto un modelo ligero de deep learning (DL) que alcanza una precisión de hasta 99.22% en la clasificación binaria y el 95.93% en tareas de clasificación múltiple, superando a modelos previos y demostrando la eficiencia de estos enfoques en la detección de la enfermedad [7].

La capacidad de estos modelos para ser entrenados y validados en múltiples fuentes de datos independientes, como los proporcionados por la Iniciativa de Neuroimagen de la Enfermedad del Alzheimer (ADNI) y el OASIS, subraya su potencial generalizabilidad y aplicabilidad en diversos contextos clínicos [8]. El avance en la automatización de la extracción de características y la clasificación mediante el uso de DL no solo promete mejorar la precisión en la detección temprana del Alzheimer, sino también facilita un seguimiento más eficiente de la progresión de la enfermedad, lo cual es esencial para el desarrollo de estrategias de manejo y tratamiento efectivo.

El uso de técnicas de ML y DL en la base de datos OASIS ha permitido abordar diversas problemáticas en el campo de

la neuroimagen. Un estudio importante de Aditya et al. [9] desarrolló una escala calibrada interpretable para medir similitudes en imágenes neurológicas. Este trabajo utilizó un modelo MAA para realizar la reducción de dimensionalidad y analizar factores de riesgo e identificadores clave. Los parámetros del modelo fueron optimizados y comparados con análisis tradicionales de clasificación, lo que permitió una evaluación más precisa de las similitudes en las imágenes.

En otro estudio, Zhang et al. [10] investigaron el uso de redes neuronales profundas para la integración y extracción de características en métodos de registro deformable de imágenes. Este estudio implementó una red de registro de deformaciones, utilizando pares de imágenes T1w del proyecto HCP-A sin cráneo, lo que demostró la efectividad de estos métodos para alinear imágenes de resonancia magnética de manera precisa. La optimización de los hiperparámetros fue crucial en este contexto para asegurar la precisión y robustez del modelo.

Además, un trabajo reciente de Liu et al. [11] exploró las diferencias de sexo en las asociaciones de estructuras cerebrales mediante modelos de ecuaciones de estimación generalizadas. Este estudio utilizó técnicas avanzadas de ML para analizar las diferencias y similitudes en las mediciones neurológicas entre sexos. La metodología permitió una comprensión más profunda de cómo las variaciones estructurales en el cerebro pueden estar influenciadas por el sexo, aportando información valiosa para futuras investigaciones en el campo de la neurociencia.

El estudio "Automatic Recognition of Mild Cognitive Impairment from MRI Images Using Expedited Convolutional Neural Networks" [12] mostró que el rendimiento más efectivo se obtuvo con una combinación de CNNs, preentrenamiento y aumento de datos. Los resultados indicaron una precisión del 90.6%, una recuperación del 92.8% y una puntuación F de 89.4%. Sin ajustes, las CNNs produjeron resultados significativamente menores, con una precisión del 74.3%, una recuperación del 74.7% y una puntuación F de 72.8%. Estas métricas evidencian la importancia del preentrenamiento y el aumento de datos para mejorar el desempeño del modelo [García].

Por otro lado, en el análisis de resonancia magnética cerebral para el diagnóstico de la enfermedad de Alzheimer [13], utilizando un sistema conjunto de redes neuronales convolucionales profundas, se reportó una precisión del 78% con una recuperación del 78% y una puntuación F de 72% en el primer modelo. En el segundo modelo, se alcanzó una precisión del 77%, una recuperación del 77% y una puntuación F de 75%. Este estudio demostró la capacidad de las redes neuronales convolucionales profundas para clasificar eficazmente las imágenes de resonancia magnética, aunque la variabilidad en las métricas sugiere la necesidad de seguir optimizando los modelos y sus parámetros.

Este estudio tiene la capacidad de desarrollar un sistema de ML que pueda identificar diferentes etapas de la EA (sin EA, muy leve, leve y moderada), que con el database de OASIS, se realice mediante exámenes físicos (CDR y MMSE) y extracción de características de MR, con el fin de mejorar el diagnóstico temprano de la enfermedad de Alzheimer mediante el uso de aprendizaje automático.

3 MATERIALES Y MÉTODOS

En esta sección, se detalla el enfoque metodológico empleado para el desarrollo del modelo de inteligencia artificial destinado a detectar niveles de Alzheimer utilizando características extraídas de imágenes RM del cerebro y algunas características que proporciona la base de datos. Se presenta un proceso estructurado que abarca desde la adquisición de datos hasta la optimización del modelo, asegurando la robustez y eficacia del sistema propuesto en el diagnóstico temprano de la enfermedad de Alzheimer.

3.1 Base de datos de imágenes de RM

El dataset OASIS (OpenAccess Series of Imaging Studies) [14], lanzado en 2007, es una serie de conjuntos de datos de resonancia magnética que está disponible públicamente para su estudio y análisis. El conjunto de datos inicial consta de una colección transversal de 416 sujetos de 18 a 96 años de edad. Cien de los sujetos incluidos, mayores de 60 años, han sido diagnosticados clínicamente con EA muy leve a moderada. Todos los sujetos son diestros e incluyen tanto hombres como mujeres. Para cada sujeto, se incluyen tres o cuatro exploraciones individuales de resonancia magnética ponderada en T1 obtenidas en sesiones de imagen únicas. El formato de los datos es NIFTI y la extensión del archivo es (.nii). Los datos de RM ofrecen detalles del cerebro y visualizan la anatomía en los tres planos: axial, sagital y coronal.

El dataset proporciona tanto un conjunto de características predeterminadas como un conjunto de imágenes. Si bien las características predeterminadas ofrecen una base estructurada para el análisis, el enfoque de este estudio radica en extraer más características en las imágenes para enriquecer aún más el conjunto de características iniciales.

3.2 Preprocesamiento de imágenes

El conjunto de datos proporcionado ya ha sido sometido a un proceso de preprocesamiento exhaustivo, que incluye la segmentación de las imágenes para clasificar el tejido cerebral en líquido cefalorraquídeo, materia gris y materia blanca. Este proceso es fundamental porque permite una comprensión detallada de la anatomía cerebral y facilita el análisis de las imágenes en diferentes contextos clínicos y de investigación.

3.3 Extracción de características

Se procedió a la extracción de características a partir de las imágenes, centrándose en los cortes axiales del cerebro. Específicamente, se calculó el promedio de los slices correspondientes a las secciones 70 a 100, seleccionados debido a su contenido informativo óptimo. Se observó que los cortes cercanos a los límites, tanto inferiores como superiores, presentaban principalmente estructuras óseas del cráneo con poca información relevante sobre la anatomía cerebral.

Para la extracción de características, se aplicaron diversas técnicas, incluyendo el análisis de coeficientes de wavelet. Estos coeficientes permiten descomponer la señal de imagen en diferentes frecuencias y escalas, proporcionando información detallada sobre la textura y los bordes de las estructuras cerebrales. La matriz de co-ocurrencia de niveles de gris (GLCM) se utilizó para analizar la distribución y la relación de intensidades de píxeles adyacentes en la imagen, lo que permite caracterizar la textura y la homogeneidad de

diferentes regiones cerebrales. Este análisis puede ser especialmente útil para detectar cambios sutiles en la textura que pueden estar asociados con ciertas patologías cerebrales. El cálculo del Gray to White Matter Ratio proporciona una medida cuantitativa de la proporción de materia gris con respecto a la materia blanca en el cerebro. Esta característica es relevante porque cambios en esta proporción pueden indicar alteraciones en la estructura y la densidad de las células cerebrales, lo que a su vez puede estar relacionado con enfermedades neurodegenerativas como la EA. Este último se determinó aprovechando la segmentación previa del conjunto de datos y la información proporcionada sobre la resolución de los voxels, permitiendo calcular los volúmenes de materia gris y blanca.

Una vez obtenidas estas características adicionales, se integraron al conjunto inicial de características proporcionadas por el dataset. Este enfoque combinado busca aprovechar tanto la información estructurada preexistente como los detalles anatómicos específicos extraídos de las imágenes, con el objetivo de enriquecer y diversificar el conjunto de características disponible para el análisis y modelado posterior.

3.4 Preprocesamiento de datos

En el conjunto de datos OASIS, la enfermedad de Alzheimer se diagnostica clínicamente y se caracteriza utilizando la escala de Evaluación Clínica de Demencia (CDR, por sus siglas en inglés). Medidas demográficas, clínicas y derivadas de imágenes (Figura 1) están disponibles en formatos XML.

Se realizó un Análisis Exploratorio de Datos (EDA) para comprender mejor la estructura y las características del conjunto de datos. Durante este proceso, se examinaron en detalle las distribuciones de las variables, las relaciones entre las características, y se identificaron posibles patrones o anomalías en los datos. El EDA también incluyó la visualización de los datos mediante gráficos estadísticos, como histogramas y diagramas de caja, lo que permitió una comprensión más profunda de la distribución y la variabilidad de las variables.

Además, se llevaron a cabo análisis específicos para explorar la relación entre las características y la variable objetivo, en este caso, el nivel de demencia según la escala CDR. Esto implicó examinar las diferencias en las distribuciones de las características entre los grupos de pacientes con y sin demencia, así como la búsqueda de posibles correlaciones entre las características y el estado de demencia.

3.5 Desarrollo del modelo

El Random Forest (RF, por sus siglas en inglés) es una potente técnica de Machine Learning de aprendizaje supervisado que se utiliza para reducir la varianza en la estimación de una función de predicción. Esta técnica es una modificación significativa del método de ensamblado conocido como bagging, el cual construye una amplia colección de árboles de decisión descorrelacionados y luego los promedia. La clave del bagging es combinar múltiples modelos independientes pero ruidosos para reducir la varianza, aprovechando la capacidad de los árboles para capturar interacciones complejas en los datos [15]. En este contexto, se optó por utilizar el Random Forest para clasificar los datos según los niveles de Alzheimer utilizando las características proporcionadas.

Age	Age at time of image acquisition (years)
Sex	Sex (male or female)
Education	Years of education
Socioeconomic status	Assessed by the Hollingshead Index of Social Position and classified into categories from 1 (<i>highest status</i>) to 5 (<i>lowest status</i>) (Hollingshead, 1957)
MMSE score	Ranges from 0 (<i>worst</i>) to 30 (<i>best</i>) (Folstein, Folstein, & McHugh, 1975)
CDR scale	0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2 = moderate AD (Morris, 1993)
Atlas scaling factor	Computed scaling factor (unitless) that transforms native-space brain and skull to the atlas target (i.e., the determinant of the transform matrix) (Buckner et al., 2004)
eTIV	Estimated total intracranial volume (cm ³) (Buckner et al., 2004)
nWBV	Expressed as the percent of all voxels in the atlas-masked image that are labeled as gray or white matter by the automated tissue segmentation process (Fotenos et al., 2005)

Fig. 1. Características iniciales del dataset OASIS

3.6 Validación del modelo

Se evaluaron métricas para medir el rendimiento del modelo, como la precisión, la sensibilidad y la especificidad. Los verdaderos positivos (TP) se refieren a las tuplas positivas del clasificador que fueron etiquetadas correctamente. Sea TP el número de verdaderos positivos. Los falsos positivos (FP) son las tuplas negativas que fueron incorrectamente etiquetadas como positivas. Sea FP el número de falsos positivos. Los verdaderos negativos (TN) son las tuplas negativas que el clasificador etiquetó correctamente. Sea TN el número de verdaderos negativos. Los falsos negativos (FN) son las tuplas positivas que fueron incorrectamente etiquetadas como negativas. Sea FN el número de falsos negativos.

3.7 Optimización del modelo

- Ajuste de los hiperparámetros del modelo.
- Realización de mejoras en el algoritmo para mejorar su precisión y generalización.

4 RESULTADOS

4.1 Necesidad de balancear datos

Dada la disparidad de datos diferentes en cada categoría, exceso de datos con no demencia (cerca al doble de baja demencia), se

desea balancear los datos con el fin de obtener mejores resultados. Para ello se hace uso de Over-sampling, de diferentes maneras, los resultados que se muestran a continuación se obtienen con las 10 características más correlacionadas, eliminando el feature MMSE, por ser una variable muy correlacionada:

1. Smoteenn Primero, SMOTE genera nuevas instancias sintéticas de la clase minoritaria, luego ENN limpia el dataset eliminando ejemplos ruidosos y redundantes, lo que resulta en un dataset más balanceado y de mejor calidad. Los resultados son los que se muestran el reporte 2

```

Results for SMOTEENN
Accuracy: 0.975609756097561
Classification Report:

```

	precision	recall	f1-score	support
0	1.00	0.93	0.97	15
1	0.90	1.00	0.95	9
2	1.00	1.00	1.00	17
accuracy			0.98	41
macro avg	0.97	0.98	0.97	41
weighted avg	0.98	0.98	0.98	41

Fig. 2. Resultados de oversampling por smoteenn

2. ADASYN Genera muestras sintéticas de la clase minoritaria de manera adaptativa, enfocándose en las instancias minoritarias más difíciles de aprender. A diferencia de SMOTE, que genera muestras sintéticas distribuidas uniformemente, ADASYN ajusta la distribución de las muestras sintéticas basándose en la dificultad de clasificación, generando más muestras en las áreas donde la clase minoritaria es más escasa. Los resultados son los que se muestran el reporte 3

```

Results for ADASYN
Accuracy: 0.7835051546391752
Classification Report:

```

	precision	recall	f1-score	support
0	0.70	0.72	0.71	29
1	0.81	0.66	0.72	38
2	0.83	1.00	0.91	30
accuracy			0.78	97
macro avg	0.78	0.79	0.78	97
weighted avg	0.78	0.78	0.78	97

Fig. 3. Resultados de oversampling por Adasyn

3. RandomOverSampler Una de las técnicas más simples de oversampling es la sobreesampleo aleatorio, donde se duplican aleatoriamente las muestras de la clase minoritaria hasta que se equilibren las clases. Los resultados son los que se muestran el reporte 4
4. SMOTE (Synthetic Minority Over-sampling Technique) SMOTE genera muestras sintéticas en lugar de duplicar las

```

Results for RandomOverSampler
Accuracy: 0.8367346938775511
Classification Report:

```

	precision	recall	f1-score	support
0	0.81	0.70	0.75	30
1	0.81	0.81	0.81	37
2	0.89	1.00	0.94	31
accuracy			0.84	98
macro avg	0.83	0.84	0.83	98
weighted avg	0.83	0.84	0.83	98

Fig. 4. Resultados de oversampling por RandomOverSampler

muestras originales. Los resultados son los que se muestran el reporte 5

```

Results for SMOTE
Accuracy: 0.6836734693877551
Classification Report:

```

	precision	recall	f1-score	support
0	0.65	0.57	0.61	30
1	0.62	0.62	0.62	37
2	0.77	0.87	0.82	31
accuracy			0.68	98
macro avg	0.68	0.69	0.68	98
weighted avg	0.68	0.68	0.68	98

Fig. 5. Resultados de oversampling por SMOTE

5. Borderline-SMOTE Borderline-SMOTE es una variación de SMOTE que se enfoca en las instancias de la clase minoritaria que están cerca de la frontera de decisión. Los resultados son los que se muestran el reporte 6

```

Results for BorderlineSMOTE
Accuracy: 0.6836734693877551
Classification Report:

```

	precision	recall	f1-score	support
0	0.67	0.60	0.63	30
1	0.64	0.62	0.63	37
2	0.74	0.84	0.79	31
accuracy			0.68	98
macro avg	0.68	0.69	0.68	98
weighted avg	0.68	0.68	0.68	98

Fig. 6. Resultados de oversampling por Borderline-SMOTE

6. SVM SMOTE SVM SMOTE utiliza un SVM para determinar la frontera de decisión y genera nuevas muestras sintéticas cerca de esta frontera. Los resultados son los que se muestran el reporte 7
7. KMeans SMOTE KMeans SMOTE aplica K-means clustering antes de realizar SMOTE para agrupar las muestras de la clase

```

Results for SVMsMOTE
Accuracy: 0.7555555555555555
Classification Report:

```

	precision	recall	f1-score	support
0	0.75	0.77	0.76	31
1	0.68	0.52	0.59	25
2	0.79	0.91	0.85	34
accuracy			0.76	90
macro avg	0.74	0.74	0.73	90
weighted avg	0.75	0.76	0.75	90

Fig. 7. Resultados de oversampling por SVMsMOTE

minoritaria y generar nuevas muestras sintéticas dentro de esos clusters. Los resultados son los que se muestran en el reporte 8

```

Results for KMeansSMOTE
Accuracy: 0.7474747474747475
Classification Report:

```

	precision	recall	f1-score	support
0	0.68	0.68	0.68	31
1	0.69	0.69	0.69	36
2	0.88	0.88	0.88	32
accuracy			0.75	99
macro avg	0.75	0.75	0.75	99
weighted avg	0.75	0.75	0.75	99

Fig. 8. Resultados de oversampling por KMeansSMOTE

5 DISCUSSION

This paper aims to answer the

6 CONCLUSION

In this paper, we aimed to understand the

7 FUTURE WORKS

The data

ACKNOWLEDGEMENT

REFERENCES

- [1] Alzheimer's Disease International. *World Alzheimer Report 2018—The State of the Art of Dementia Research: New Frontiers*, volume 48. Alzheimer's Disease International, London, UK, 2018.
- [2] Archibaldo Donoso. La enfermedad de alzheimer. *Revista chilena de neuro-psiquiatría*, 41:13–22, 2003.
- [3] M Romano, Maria Daniela Nissen, N Del Huerto, and C Parquet. Enfermedad de alzheimer. *Revista de posgrado de la vía cátedra de medicina*, 75:9–12, 2007.
- [4] Myriam Barandiaran. Neuropsicología y diagnóstico temprano. *Revista Española de Geriatría y Gerontología*, 46:42–46, 2011.
- [5] Jianpan Huang, Peter C M van Zijl, Xiongqi Han, Celia M Dong, Gerald W Y Cheng, Kai-Hei Tse, Linda Knutsson, Lin Chen, Joseph H C Lai, Ed X Wu, Jiadi Xu, and Kannie W Y Chan. Altered d-glucose in brain parenchyma and cerebrospinal fluid of early alzheimer's disease detected by dynamic glucose-enhanced MRI. *Sci. Adv.*, 6(20):eaba3884, May 2020.
- [6] Duaa AlSaeed and Samar Fouad Omar. Brain mri analysis for alzheimer's disease diagnosis using cnn-based feature extraction and machine learning. *Sensors*, 22(8), 2022.
- [7] Ahmed A. Abd El-Latif, Samia Allaoua Chelloug, Maali Alabdulhafith, and Mohamed Hammad. Accurate detection of alzheimer's disease using lightweight deep learning model on mri data. *Diagnostics*, 13(7), 2023.
- [8] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507, 2007.
- [9] C.R. Aditya and M.B. Sanjay Pande. Devising an interpretable calibrated scale to quantitatively assess the dementia stage of subjects with alzheimer's disease: A machine learning approach. *Informatics in Medicine Unlocked*, 6:28–35, 2017.
- [10] Liutong Zhang, Guochen Ning, Lei Zhou, and Hongen Liao. Symmetric pyramid network for medical image inverse consistent diffeomorphic registration. *Computerized Medical Imaging and Graphics*, 104:102184, 2023.
- [11] Sophia Z Liu, Ghazal Tahmasebi, Ying Sheng, Ivo D Dinov, Dennis Tsilimingras, and Xuefeng Liu. Sex difference in the associations of socioeconomic status, cognitive function and brain volume with dementia in old adults: Findings from the oasis study. *medRxiv*, pages 2023–01, 2023.
- [12] Shuqiang Wang, Yanyan Shen, Wei Chen, Tengfei Xiao, and Jinxing Hu. Automatic recognition of mild cognitive impairment from mri images using expedited convolutional neural networks. In Alessandra Lintas, Stefano Rovetta, Paul F.M.J. Verschure, and Alessandro E.P. Villa, editors, *Artificial Neural Networks and Machine Learning – ICANN 2017*, pages 373–380, Cham, 2017. Springer International Publishing.
- [13] Jyoti Islam and Yanqing Zhang. Brain mri analysis for alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain informatics*, 5:1–14, 2018.
- [14] Daniel S. Marcus, Tracy H. Wang, Jamie Parker, John G. Csernansky, John C. Morris, and Randy L. Buckner. Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, September 2007.
- [15] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*. Springer series in statistics. Springer, New York, NY, 2 edition, February 2009.