

RESEARCH ARTICLE OPEN ACCESS

A Machine-Learning Classification for Torlesse Composite Terrane Petrofacies and its Application to Sediment Provenance

Matthew O. Parker¹  | David E. Dempsey²  | Matthew W. Sagar³  | Kari N. Bassett¹  | Alexander R. L. Nichols¹  | Greg H. Browne³ 

¹School of Earth and Environment, University of Canterbury, Christchurch, New Zealand | ²Department of Civil and Environmental Engineering, University of Canterbury, Christchurch, New Zealand | ³Earth Sciences New Zealand, Lower Hutt, New Zealand

Correspondence: Matthew O. Parker (matt.parker@pg.canterbury.ac.nz)

Received: 30 September 2025 | **Revised:** 8 May 2026 | **Accepted:** 11 May 2026

Keywords: geochemistry | Kaweka Terrane | machine learning | Pahau Terrane | provenance | Rakaia Terrane | Torlesse Composite Terrane

ABSTRACT

The Torlesse Composite Terrane (TCT) forms many of the mountain ranges in Aotearoa New Zealand and has provided enormous quantities of coarse-grained sediment to Te Riu-a-Māui/Zealandia's basins since the mid-Cretaceous. Tracing the provenance of these sediments to certain regions of the TCT can indirectly reconstruct exhumation patterns associated with the development of the Australian–Pacific plate boundary. Supplemented by new analyses, we are now able to include one of the new constituent terranes of the TCT (the Kaweka Terrane), which had not been recognised when the previous geochemical classification was developed. We train four different machine learning algorithms using samples from Te Waipounamu/the South Island to classify TCT-derived sandstone conglomerate clasts into four different petrofacies/terrane of the TCT using log-transformed trace element ratios. Balanced accuracies for the four algorithms used are 77% (linear support vector machine), 69% (k-nearest neighbours), 75% (random forest), and 77% (neural network). This rises to 80% when an equal voting system is used. These accuracies are similar to existing classifications for the TCT but include the Kaweka Terrane as a classification option. We then show an example of how this updated classification can be applied to late Cenozoic TCT-derived conglomerate clasts.

1 | Introduction

The Torlesse Composite Terrane (TCT) is the largest basement unit in Aotearoa New Zealand by exposed area (Mortimer 2004) and forms much of Kā Tiritiri o te Moana/the Southern Alps and the axial ranges of Te Ika-a-Māui/the North Island (Figure 1). A small outlier also occurs at the Chatham Islands (Adams et al. 2008). From mid-Cretaceous to Recent time, the TCT has supplied vast quantities of sediment to several of Te Riu-a-Māui/Zealandia's basins (e.g., Crampton and Landis 1988; Field and Browne 1989; Smale 1990; Mortimer et al. 2001; Browne and Naish 2003; Cutten et al. 2006; Adams et al. 2013; Adams et al. 2017; Rotzien et al. 2018; Gooley and

Nieminski 2022; Parker 2022; Sagar et al. 2024). For provenance studies, especially those focusing on late Cenozoic conglomerates that signal the exhumation of basement rocks, it becomes desirable to attempt to resolve their sources to more specific areas within the TCT.

While the TCT is well-known for its lithological monotony of sandstone and mudstone (e.g., Mortimer 2004), there are subtle spatial variations in modal mineralogy (MacKinnon 1980, 1983), which are also reflected in major and trace element whole-rock X-ray fluorescence (XRF) geochemistry (Roser et al. 1995; Roser and Korsch 1999). However, some overlap remains between different regions of the TCT using the existing geochemical

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *New Zealand Journal of Geology and Geophysics* published by John Wiley & Sons Australia, Ltd on behalf of Royal Society of New Zealand Te Apārangi.

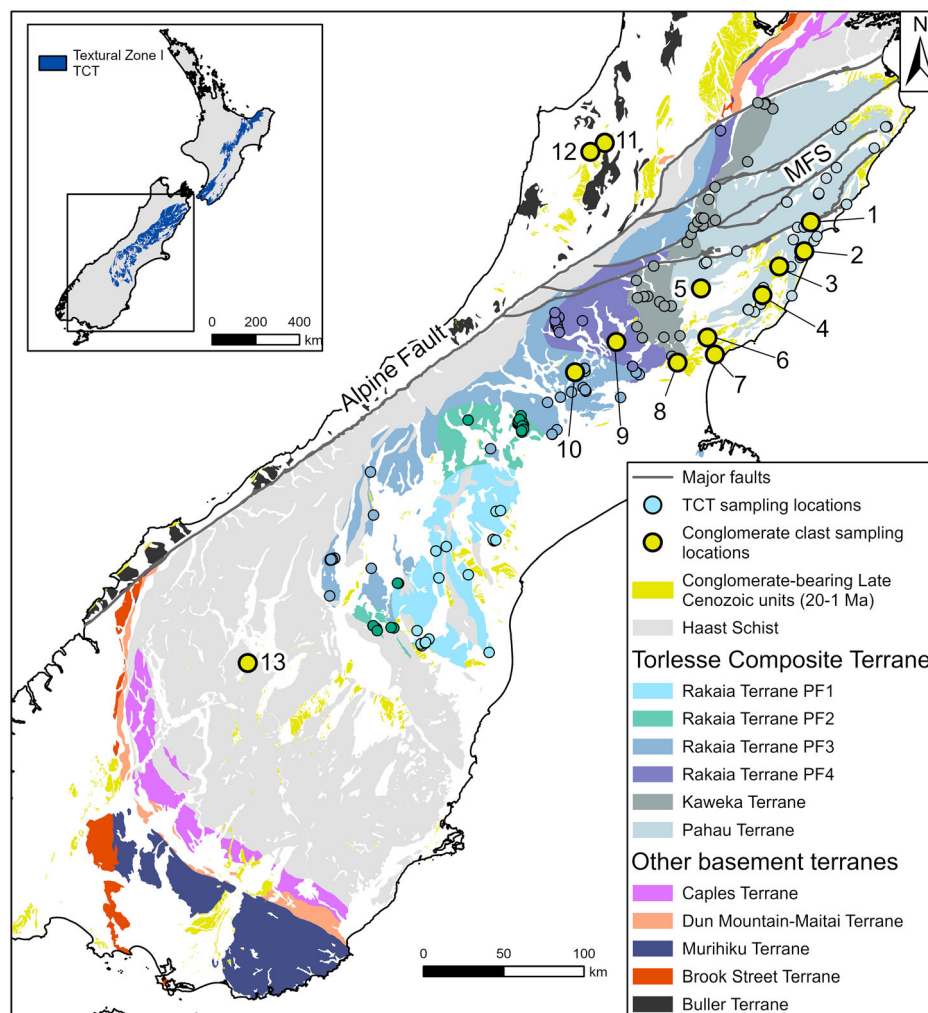


FIGURE 1 | Basement geological map of Te Waipounamu/the South Island showing relevant sandstone-bearing basement units, the spatial extents of the terranes/petrofacies that form the Torlesse Composite Terrane and late Cenozoic conglomerate-bearing units. The locations of 200 Textural Zone I sandstones samples used in this classification that were analysed for whole rock XRF major and trace element geochemistry and sandstone conglomerate sampling locations are also shown. Conglomerate sampling locations 1–10 are from [Parker \(2022\)](#), 11–12 are from [Mortimer et al. \(2001\)](#), and 13 is from [McDonnell and Craw \(2003\)](#). Geology modified from [Heron \(2020\)](#). Petrofacies from [MacKinnon \(1980\)](#) and [Wandres \(2002\)](#).

classification diagram ([Roser and Korsch 1999](#)). The current classification also has eye-drawn boundaries and was unable to include a now constituent terrane within the TCT as it had not yet been identified when the classification was published.

As a result of this, a new classification method that can include the new terrane, maintain/improve classification accuracy, reduce bias, and utilise several relatively immobile element ratios is necessary. The use of supervised machine learning algorithms is one strategy that can fulfil these requirements. Machine learning algorithms, including support vector machine, random forest, decision tree, logistic regression, nearest neighbours, neural networks (such as convolutional neural networks, multi-layer perceptrons, and transformers), and sparse multinomial regression, have successfully been trained using geochemical variables and have been applied previously to large datasets (e.g., [Petrelli and Perugini 2016](#); [Petrelli et al. 2017](#); [Ueki et al. 2018](#); [Lui et al. 2022](#); [Tamanna et al. 2025](#)). Generally, high (>80%) out-of-sample accuracies are commonly reported in these studies,

indicating that these methods can be applied successfully to geochemical datasets.

This paper presents an updated classification for the TCT in Te Waipounamu/the South Island using multiple machine learning algorithms. These models are trained on existing and new TCT data to predict the assignment of sandstone conglomerate clasts to specific regions within the TCT. All datasets and the python code used in this dataset are available as supplementary material.

2 | Background

2.1 | Artificial Intelligence in the Geosciences

Over the past decade or so, there has been an increasing number of geoscience studies that have developed, tested, or utilised workflows that involve artificial intelligence (e.g., [Fleming et al. 2021](#); [Lachowycz 2024](#)). These workflows help to reduce any operator biases while improving consistency, reproducibility,

and efficiency of classification tasks when compared to manual methods. Most artificial intelligence applications in geoscience studies generally use supervised machine learning, where datasets contain expert-assigned labels (e.g., rock type, facies) prior to training a classification model.

Recent research has highlighted how machine and deep learning can be applied to lithofacies classifications of rock core data using wireline logs (Mandal and Rezaee 2019; Wood 2021) or automated semantic segregation performed on high-resolution core images (Di Martino et al. 2023). Training supervised machine-learning algorithms using major and/or trace element geochemical data has also been widely applied in the geosciences, for example, discriminating the tectonic setting of igneous rocks (Petrelli and Perugini 2016; Ueki et al. 2018; Doucet et al. 2022; Ou et al. 2024), locating tephra sources (Petrelli et al. 2017), fingerprinting quarry source locations (Dornan et al. 2020), predicting soil sample geochemistry (Lui et al. 2022), deriving the mineral-based brittleness index in rock cores (Hassan et al. 2022) and performing automatic rock classification using the well-known total alkali-silica (TAS) plot (Tamanna et al. 2025).

2.2 | Stratigraphy and Lithologies of the Torlesse Composite Terrane

The TCT is part of the Cambrian–mid Cretaceous basement (Austral Superprovince) of Te Riu-a-Maui/Zealandia that formed part of eastern Gondwana adjacent to Australia and West Antarctica before the initiation of sea floor spreading in the Late Cretaceous (Gaina et al. 1998; Mortimer et al. 2014; Riefstahl et al. 2020).

The TCT comprises three main tectonostratigraphic units, the Rakaia (Carboniferous–Late Triassic), Kaweka (Early–Middle Jurassic), and Pahau (Early Cretaceous) terranes (Mortimer et al. 2014; Mortimer 2025). In addition to more detailed lithostratigraphic subdivisions (e.g., Mount Taylor Group, Mount Potts Group, Pahau River Group), the TCT has also been divided into fossil (or faunal) zones, with up to 20 previously recognised (Campbell and Warren 1965; Landis and Bishop 1972; Andrews et al. 1976; MacKinnon 1980). However, many of these are represented by very few fossil occurrences, or are not geographically exclusive (i.e., overlap with other zones), which warranted the simplified subdivision into five geographically extensive fossil zones (MacKinnon 1980). Each of these contains many age diagnostic fossil localities, with almost none located outside their respective zone boundaries (MacKinnon 1983). The ages for the fossil zones are as follows: 1 – Permian, 2 – Middle Triassic, 3 – late Middle Triassic–early Late Triassic, 4 – Late Triassic, 5 – Late Jurassic–Early Cretaceous (MacKinnon 1983). Fossil zones 1–4 are restricted to the Rakaia Terrane and regionally young towards the northeast (toward the boundary with the Kaweka Terrane). Fossil zone 5 occurs only within the Pahau Terrane and small regions containing Jurassic shallow marine/terrestrial strata enclosed by the Rakaia Terrane (MacKinnon 1980). The Kaweka Terrane was not originally classified as a separate fossil zone and has been previously included in neighbouring fossil zones (Landis and Bishop 1972; Andrews et al. 1976) and is the boundary between fossil zones 4 and 5 of MacKinnon (1980). It is now recognised as a terrane from a distinctive Early Jurassic peak in detrital zircon age spectra in both the North and South

islands (Adams et al. 2011). There is a general younging direction towards the northeast, and the younger petrofacies/terrane are offset by faults within the Marlborough Fault System (Figure 1) (e.g., Rattenbury et al. 2006).

Quartzofeldspathic sandstone and mudstone (commonly referred to in Aotearoa New Zealand as greywacke and argillite) dominate the TCT, with rare but widespread conglomerate, red and green mudstone, chert, limestone, and basalt (Bradshaw 1972; Andrews et al. 1976; Smale 1978; MacKinnon 1983). Sandstone and mudstone are also widespread within other basement terranes in Aotearoa New Zealand (e.g., Buller, Brook Street, Murihiku, Caples, and Dun Mountain–Maitai terranes) (Figure 1) but generally have slightly to substantially different modal mineralogy and whole-rock geochemistry (Roser and Korsch 1988; Roser et al. 1996; Roser and Korsch 1999; Roser et al. 2002). Zones of *mélange* (sheared sandstone and mudstone with exotic blocks of chert, basalt, and limestone) are also present within the TCT (Bradshaw 1973; Botsford 1983; Silberling et al. 1988; Forsyth et al. 2008). A prominent zone of this lithology, known as the Esk Head Belt, is included as part of the Kaweka Terrane (Botsford 1983; Adams et al. 2011). Much of the TCT is low-grade (zeolite or prehnite–pumpellyite) metamorphic facies or (Textural Zone (TZ) I of Turnbull et al. 2001) and can still be considered sandstone (Botsford 1983; Rattenbury et al. 2006; Forsyth et al. 2008). Higher-grade metamorphosed equivalents of the TCT (TZII–IV) (included in the Haast Schist) have been exhumed in Otago, Marlborough, and along a narrow band adjacent to the Alpine Fault (e.g., Turnbull et al. 2001; Cooper and Palin 2018; Mortimer 2025).

2.3 | Compositional Variation in the Torlesse Composite Terrane

MacKinnon (1980) sampled sandstone throughout each of the five major fossil zones and noted subtle but consistent changes in modal mineralogy. As a result, MacKinnon (1980) introduced five fossil-defined petrofacies (PF1–5) (Figure 1), which are mappable regions of a lithostratigraphic unit with similar composition and age (Dickinson and Rich 1972; MacKinnon 1980; Mortimer 1994). The petrofacies share the same boundaries and numbering system as the fossil zones.

These changes are well displayed using the quartz–feldspar–lithics (QFL) ternary plot (Figure 2), where the geometric means of each petrofacies show small temporal compositional changes. The proportion of lithics decreases between Rakaia Terrane PF1–3, followed by an increase in quartz and lithic proportions in Rakaia Terrane PF4, and in the Pahau Terrane, lithic proportions exceed those in Rakaia Terrane PF1 (MacKinnon 1983). Other studies in the South Island confirm similar results to the Rakaia Terrane petrofacies of MacKinnon (1980) (e.g., Dickinson 1971; Beggs 1980; Hicks 1981). However, studies on the Pahau Terrane in the South Island (Crampton and Landis 1988; Orłowski 2001) note lower lithic proportions than those of MacKinnon (1980) and overlap with Rakaia Terrane PF4. As the Kaweka Terrane was not included in the MacKinnon (1980) petrofacies classification, only six sandstone samples have been point-counted, all from the Esk Head–Okuku area (Botsford 1983), and may not be representative of the whole terrane. These

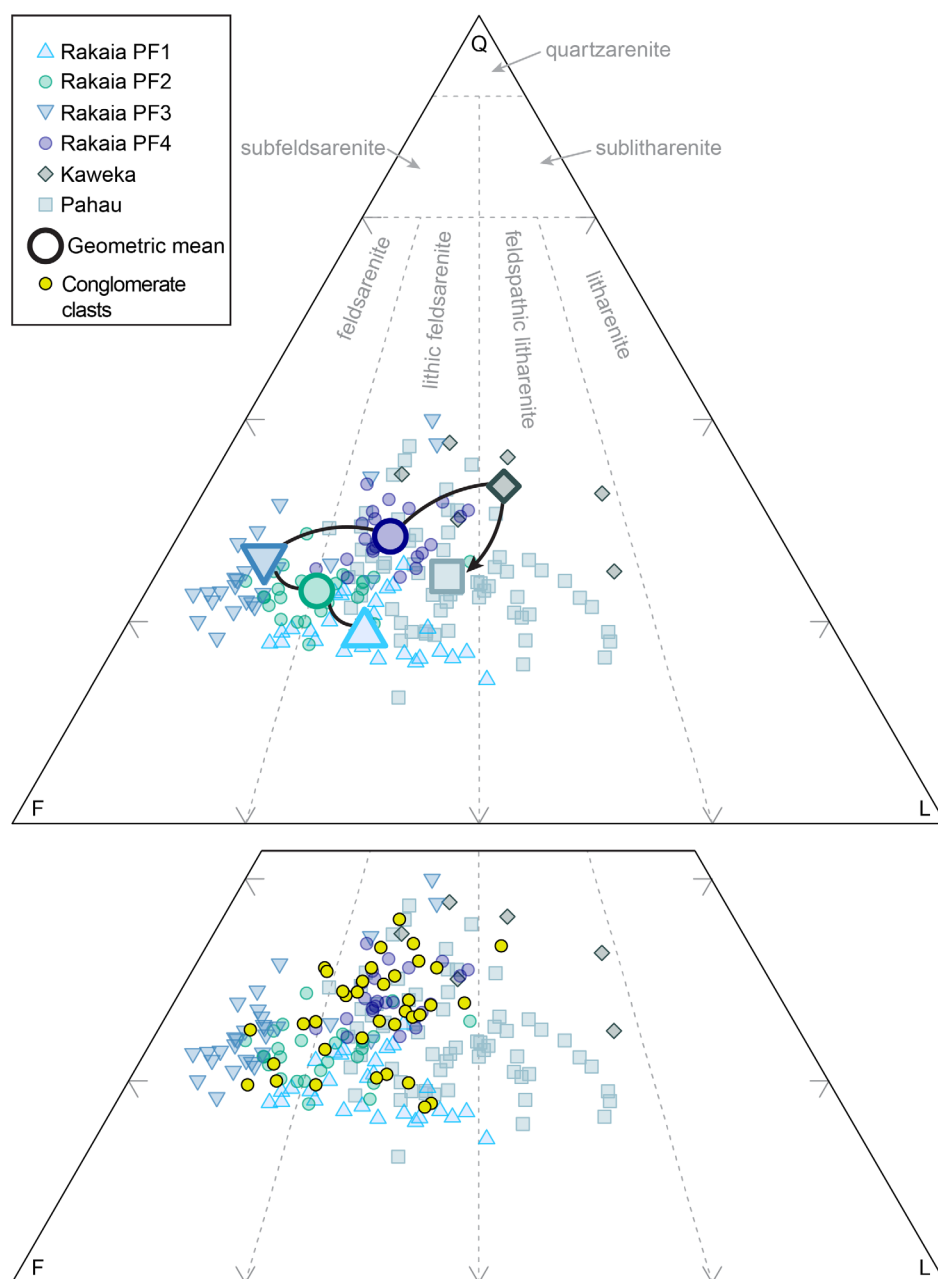


FIGURE 2 | Top: Quartz–feldspar–lithics (QFL) ternary plot with quartz–feldspar–rock fragments (QFR) classification (Folk et al. 1970) showing point count data for sandstone samples from each of the petrofacies/terranes of the Torlesse Composite Terrane (TCT). Bottom: TCT QFL data with late Cenozoic sandstone clasts overlain. TCT Data from Dickinson (1971), Beggs (1980), MacKinnon (1980), Botsford (1983), Crampton and Landis (1988) and Orlowski (2001). Conglomerate clast data from McDonnell and Craw (2003) and Parker (2022). Staining for feldspar was used in almost all previous studies. The exceptions are Hicks (1981), where there is no mention of staining, and Parker (2022), in which staining was unsuccessful and quartz and feldspar were differentiated based on the presence/absence of undulatory extinction and/or interference colours.

samples generally contain less feldspar than the Rakaia and Pahau terranes, however Botsford (1983) appeared to have used a different point-counting technique as their results are shown on a quartz–feldspar–rock fragments (QFR) plot. Most other studies used the Gazzi–Dickinson technique, or a very minor modification to the Gazzi–Dickinson technique (see MacKinnon 1980).

The geochemical validity of MacKinnon (1980) TCT petrofacies was confirmed by Roser and Korsch (1999), who noticed subtle variations using major and trace element geochemistry of the same samples that MacKinnon (1980) point-counted and

classified into petrofacies, with additional samples from fossil zones in both the North and South islands (see Roser et al. 1995). Despite noted modal mineralogical variability in the South Island Pahau Terrane (Crampton and Landis 1988; Orlowski 2001) when compared to the data of MacKinnon (1980), the geochemistry of these samples from all authors is similar (Roser et al. 1995).

Roser and Korsch (1999) utilised several trace element ratios that used felsic- (Th, La, Ce, La, Zr) and mafic-linked (Sc, Y, V, Ti) elements to create maximum differences. Mean ratios of Th/Sc,

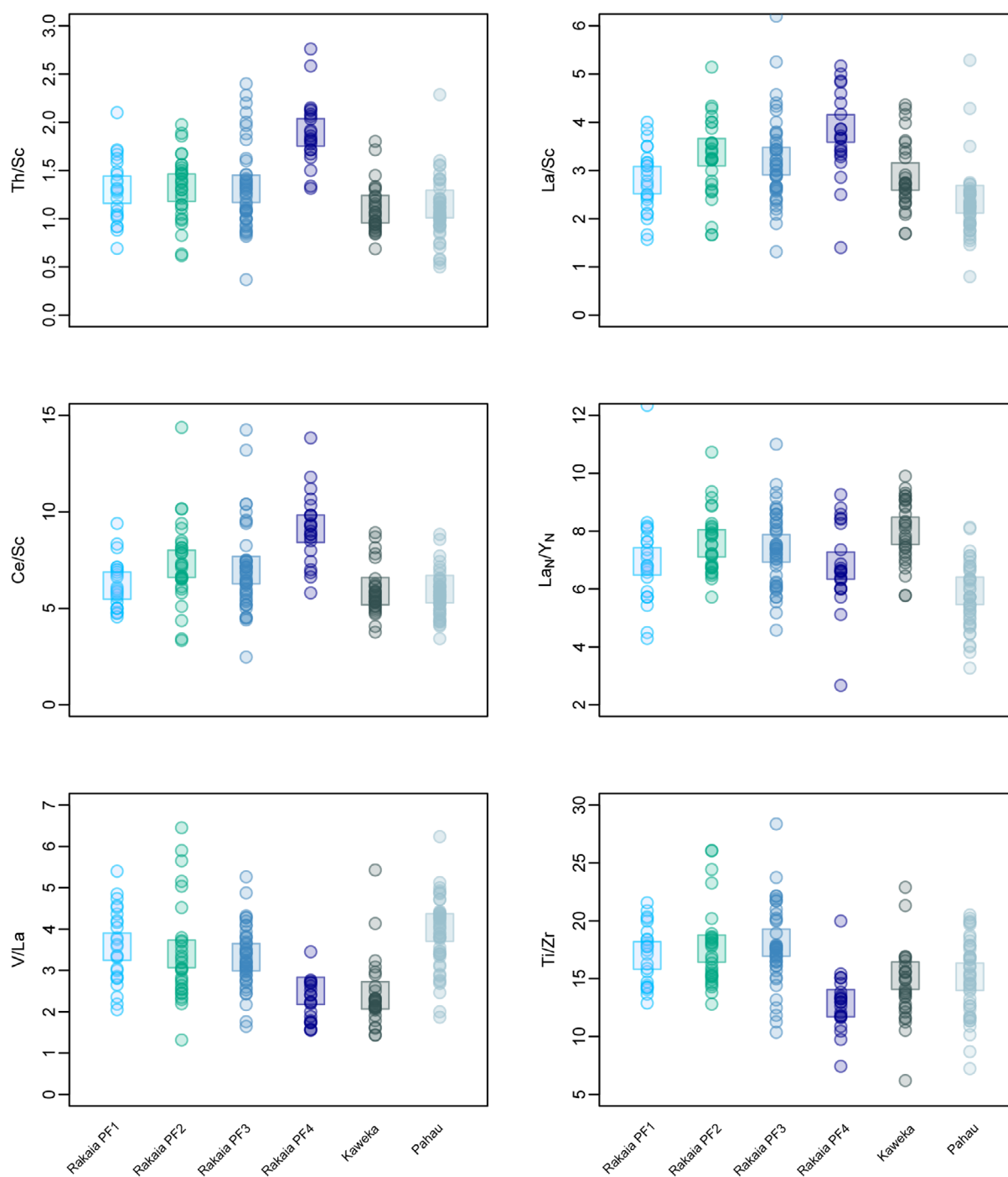


FIGURE 3 | Trace element ratios of Roser and Korsch (1999) for the different petrofacies/terraces of the TCT. Only the 200 Torlesse Composite Terrane (TCT) sandstone samples used in our classification are shown. Data from Roser et al. (1995), Wandres (2002), the Petlab database (Strong et al. 2016) and this study. The La_N/Y_N ratio uses chondritic rare earth element normalising factors from Taylor and McLennan (1985).

La/Sc, and Ce/Sc are similar/slightly increase between Rakaia Terrane PF1–3, increase sharply to Rakaia Terrane PF4, before sharply decreasing to the Kaweka and Pahau terranes (Figure 3). Increases in these ratios signify original source evolution and/or increasing sedimentary maturity (Roser and Korsch 1999). The La_N/Y_N ratio represents the slope of rare earth element patterns (Roser and Korsch 1999) and uses chondritic rare earth element normalising factors of Taylor and McLennan (1985). On average, La_N/Y_N increases steadily from Rakaia Terrane PF1–3, decreases in Rakaia Terrane PF4, increases to a maximum value in the Kaweka

Terrane, before dropping sharply to the minimum average value in the Pahau Terrane (Figure 3). The inverse ratios (Ti/Zr, V/La) generally mirror Th/Sc, La/Sc, and Ce/Sc; however, Rakaia Terrane PF4 and the Kaweka Terrane have similar V/La ratios (Figure 3).

2.4 | Existing Geochemical Classifications for the TCT

Currently, the main method of classifying TCT sandstone into specific petrofacies and/or terranes is to visually assess sample

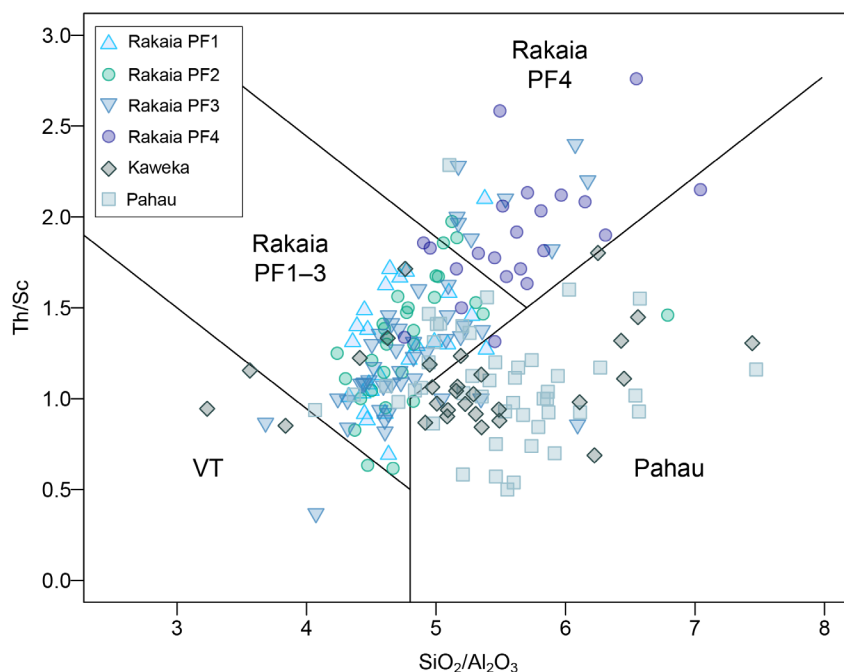


FIGURE 4 | Th/Sc–SiO₂/Al₂O₃ classification plot from Roser and Korsch (1999). VT = volcanogenic terrane (including Brook Street, Dun Mountain–Maitai, Caples, and Waipapa terranes). Only the 200 Torlesse Composite Terrane sandstone samples used in our classification are shown. Data sources as for Figure 3.

location(s) on the Th/Sc–SiO₂/Al₂O₃ plot (Figure 4) (Roser and Korsch 1999). The Th/Sc ratio utilises two relatively immobile trace elements which retain source rock characteristics and are expected to increase with sediment maturation via recycling and/or source rock petrologic evolution (Roser and Korsch 1999). The SiO₂/Al₂O₃ ratio utilises major elements with differing mobility and is a proxy for sediment maturity; values increase as quartz is concentrated during weathering and/or recycling (Roser and Korsch 1986). This plot was initially used to help model the derivation of Rakaia Terrane PF4 and the Pahau Terrane from the Rakaia Terrane PF1–3 and more volcanogenic Eastern Province terranes (e.g., Caples and Dun Mountain–Maitai terranes) with boundaries drawn by eye (Roser and Korsch 1999). Rakaia Terrane PF1–3 is grouped as these petrofacies have relatively uniform major and trace elements and trace element ratios, whereas Rakaia Terrane PF4 and the Pahau Terrane are more distinct (Roser and Korsch 1999). As the Kaweka Terrane was thought to be a zone of mélange between the Rakaia and Pahau terranes at this point, it was not included in the classification plot.

Linear discriminant analysis (LDA) has previously been used to discriminate between TCT petrofacies/terrane and classify new samples (Roser et al. 1993; Parker 2022). This method calculates a set of linear functions based on multiple input variables that find maximum separation between group means for groups that are pre-defined (Boedeker and Kearns 2019). This has been used in the southern North Island, where the Rakaia Terrane PF3, Rakaia Terrane PF4, and the Pahau Terrane were discriminated using samples from the Wellington Peninsula and the Aorangi Range using 15 elements (not listed) in Roser et al. (1993). This classification is unpublished but has been used to classify sandstone samples from the Wellington area in Begg and Mazengarb (1996, Appendix 2).

Parker (2022) also used LDA to classify TCT-derived sandstone clasts into either Rakaia Terrane PF1–3, Rakaia Terrane PF4, or the Pahau Terrane in northern Canterbury. This classification used transformed natural log-ratios of most of the trace elements used by Roser and Korsch (1999) (Ce, La, Sc, Th, V, and Zr) with a common denominator (Ti). This is known as additive log-ratio transformation (Aitchison 1986) and helps solve issues related to the constant sum effect, where when one element increases, other elements decrease, as geochemical data is commonly reported as weight percent (wt%) or parts per million (ppm). Yttrium was not used in this classification as it was noted that some weathered sandstone conglomerate clasts had higher La_N/Y_N values than TCT samples, whereas La/Sc values were similar. This may have indicated that this element was not completely immobile for an unknown reason (Parker 2022) or was an analytical precision issue due to low Y content (<20–30 ppm).

3 | Materials and Methods

3.1 | Classification Training Dataset

The classification uses a filtered dataset of 200 South Island Torlesse Composite Terrane sandstone samples that have been analysed for whole rock major and trace element X-ray fluorescence (XRF) geochemistry. Most geochemical analyses are from Roser et al. (1995), which was the same dataset later used for the SiO₂/Al₂O₃–Th/Sc classification (Roser and Korsch 1999), with additional samples from Wandres (2002), the Petlab database (Strong et al. 2016), and new Kaweka Terrane data acquired for this study. All samples used are Textural Zone (TZ) I sandstones and exclude the Dansey's Pass TZIIA semischist dataset (51 samples), although no apparent geochemical changes between TZI and TZII Rakaia Terrane PF2 samples are

recognisable (Roser et al. 1995; Roser and Korsch 1999). However it is noted that they are geographically distant from fossil control (MacKinnon 1980). Analytical results that were affected by probable calcite veins or cement (recognised by $\text{CaO} > 5 \text{ wt\%}$) were removed from the dataset. Parker (2022) included the Dansey's Pass TZIIA samples and samples with $> 5 \text{ wt\%}$ CaO that are excluded here in his LDA ($n = 213$).

Prior to this study, only two sandstone samples from the South Island Kaweka Terrane had been analysed and were available in the Petlab database (JKS10 and LTHM1). For this reason, a further 33 South Island Kaweka Terrane sandstone samples from the National Petrology Reference Collection were analysed for major and trace elements by XRF at Spectrachem Analytical (Verum Group Ltd), Christchurch. Major and trace elements were measured using fused glass beads and pressed powder briquettes, respectively. Only 25 of these new samples were used for the new classification due to two samples having Sc below detection limits (2 ppm), and six samples with $\text{CaO} > 5 \text{ wt\%}$. All training samples in our dataset have undergone identical data handling procedures, and all major and trace elements have been normalised to 100 wt% on an anhydrous basis.

The elements selected for this classification are identical to those selected by Parker (2022) (Ce, La, Sc, Th, Zr, and Ti), with the exception of V, which was removed as it is moderately soluble and not generally considered to be an immobile element. The values for each of the 200 samples of the training/testing/validation dataset were compiled into a single spreadsheet with their terrane/petrofacies assignment value (Rakaia PF1–3 = 1, Rakaia PF4 = 4, Kaweka = 6, Pahau = 5). New sandstone clast samples were compiled into a separate spreadsheet, with sample ID in place of the terrane assignment, assess prediction results.

3.2 | Conglomerate Clast Dataset

We have compiled all available whole-rock geochemical analyses of Middle Miocene–Early Pleistocene TZI sandstone clasts across the South Island (Figure 1). These include 27 middle Miocene–Early Pleistocene clasts from the Waima, Brechin, Kowai, and Greta formations in North Canterbury (locations 1–10 in Figure 1) (Parker 2022), six Late Pliocene–Early Pleistocene clasts from the Old Man Group in the Inangahua Valley (locations 11 and 12 in Figure 1) (Mortimer et al. 2001) and five Pliocene clasts from the Waiorau Conglomerate in the Cardrona Valley (location 13 in Figure 1) (McDonnell and Craw 2003). Conglomerates clasts from the Old Man Group, Waiorau Conglomerate, and Brechin Formation are non-marine and were deposited as part of a braided river paleoenvironment (Mortimer et al. 2001; McDonnell and Craw 2003; Parker 2022). The clasts from the Kowai Formation were deposited in a distal alluvial fan through to nearshore marine paleoenvironment, while clasts from the Greta and Waima formations were deposited on the shallow shelf (locally non-marine) through to a bathyal submarine canyon paleoenvironment (Parker 2022).

Sandstone clasts from Parker (2022) that were the least weathered, cobble sized or larger and lacked significant quartz veins were preferentially selected for geochemical analysis, but in terms of other clast attributes (e.g., rounding, shape) clast

selection was random. Weathering of sandstone conglomerate clasts is not explicitly mentioned in Mortimer et al. (2001), but many clasts are noted to be orange–brown in colour due to oxidative alteration of clast rims and clay alteration throughout the clast interiors (McDonnell and Craw 2003; Parker 2022), compared to the generally grey coloured in-situ TCT sandstone. Despite this, all sandstone conglomerate clasts are well within the limits of the TCT training data using the chemical index of alteration (CIA), plagioclase index of alteration (PIA), and mafic–felsic–weathering (MFW) plots (Figures S1 and S2). Sandstone clasts from the non-marine and nearshore units were generally collected from conglomerate-dominated intervals with thinner sandstone intervals, whereas clasts from shallow shelf–bathyal units were collected from thinner conglomerate lenses/packages within thicker mudstone and/or sandstone-dominated intervals.

In their respective studies, these clasts have been confirmed to have a generic TCT source through petrography and/or geochemistry (Figure 2), where they can be differentiated from adjacent sedimentary terranes (e.g., Brook Street, Caples, Buller, Dun-Mountain–Maitai terranes). Following normalisation of all major and trace element data to 100 wt% on an anhydrous basis, we placed the La, Sc, Th, Zr, and Ti (all ppm) values in a separate spreadsheet to the training dataset and ran the classification model.

3.3 | Model Training Pipeline

Initially, log-ratios of Ce, La, Sc, Th, V, and Zr with a Ti denominator were calculated for the datasets following Parker (2022). To reduce unnecessary information and optimise out of sample accuracy, we tested feature dropout by removing the least important elements using the random forest algorithm as a guide and reran our model until accuracies for all four algorithms decreased. We observed average balanced accuracy and Matthew's Correlation Coefficient increase slightly when $\ln(\text{V}/\text{Ti})$ and $\ln(\text{Ce}/\text{Ti})$, the two least important ratios, respectively, were removed one after the other. When the next least important ratio, $\ln(\text{Zr}/\text{Ti})$, was removed, balanced accuracies across three of the four algorithms decreased. Therefore, the final elements used were La, Sc, Th, and Zr with a Ti denominator, which are generally considered to be relatively immobile during weathering/alteration. These formed the four columns of an initial feature matrix, where each row represented one sample. The target label for supervised learning is the known classification of the sample.

Individual features were first standardised to a unit normal distribution using Scikit-learn's (Python) StandardScaler. Principal components analysis (PCA) was applied to the dataset to test if there was an advantage of reduced dimensionality; however, as the improvement was only slight, we did not persist with this approach. We also tried calculating polynomial features, but these, too, yielded no appreciable gain.

The dataset was divided into a training-validation set to tune the algorithms (70%) and an external testing set to test the classification accuracy on unseen samples (30%). Due to group size imbalance in the train-validation set (Rakaia PF1–3 $n = 66$,

Rakaia PF4 $n = 17$, Kaweka $n = 22$, Pahau $n = 36$), the three largest groups were randomly undersampled to match the number of samples in the smallest group (Rakaia PF4). Therefore, the total training-validation samples used in this classification were 68.

Four classification models from the Scikit-learn machine learning library were tested: linear support vector machine (LSVM), k-nearest neighbours (KNN), random forest (RF), and neural network (NN). The NN algorithm used a fully connected multi-layer perceptron model consisting of a single hidden layer, tested with 10, 20, and 50 neurons for a maximum of 1000 iterations, and logistic, hyperbolic tangent, and rectified linear unit activation functions were evaluated during hyperparameter tuning. These four algorithms were selected as they are widely used in similar geochemical machine learning studies and have distinct theoretical methods and assumptions. Multiple algorithms allow a voting-based confidence system based on the level of agreement amongst their predictions. These algorithms also have weaker distributional assumptions and the capacity for non-linearity, when compared to LDA. Linear boundaries between classes are also enforced by LDA, which may be inappropriate if source-mixing of element interactions imposes non-linear effects.

Each algorithm was trained using a ShuffleSplit cross-validation strategy, which involved randomly splitting the dataset into a training (75%) and validation (25%) dataset five times. At each of the five splits, a grid search was undertaken to test different hyperparameters and average balanced accuracy across all splits was recorded. The hyperparameter combination with the highest balanced accuracy is then used for prediction of samples in the external test set and new unknown samples.

4 | Classification Performance Results

The test dataset used to evaluate classification accuracy on unseen data originally comprised 30% of the full dataset, which was further supplemented with additional samples that were discarded during training rebalancing and that were not used for model training. This increased the number of test samples to 132, and 66% of the full dataset.

The LSVM results can be plotted in two dimensions as a decision map using the two most important predictors, $\ln(\text{La}/\text{Ti})$ and $\ln(\text{Th}/\text{Ti})$, which shows a projection of some of the prediction boundaries alongside the external testing dataset samples (Figure 5). These two predictors are the same as the two most important features selected during training of the RF algorithm (Figure S3).

Each of the four algorithms was used to classify the external test samples, and the classification results are displayed visually as confusion matrices and numerically as classification statistics in Figure 6. These statistics can be used to evaluate how accurate each algorithm was at classifying the test dataset and identify consistent misclassifications where the algorithm(s) are performing poorly. At a high level, there is relatively little variation between the performance of the four algorithms, with balanced accuracies varying from 69% to 77%. Matthew's Correlation Coefficient values for the four algorithms vary from 0.42 to 0.56, where a value of -1 represents no correct predictions, 0 represents random predictions, and 1 represents perfect predictions.

The petrofacies/terrane with the highest recall values (proportion of samples from a specific group that is correctly

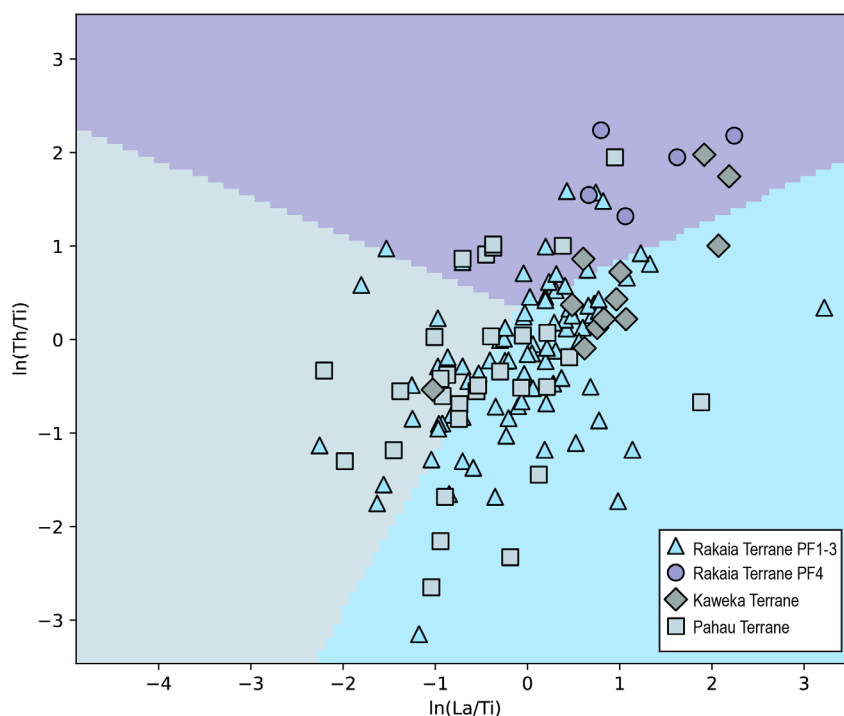


FIGURE 5 | Linear support vector machine decision plot with the two most important log-transformed ratios showing the decision boundaries for all four classification options and locations of validation samples.

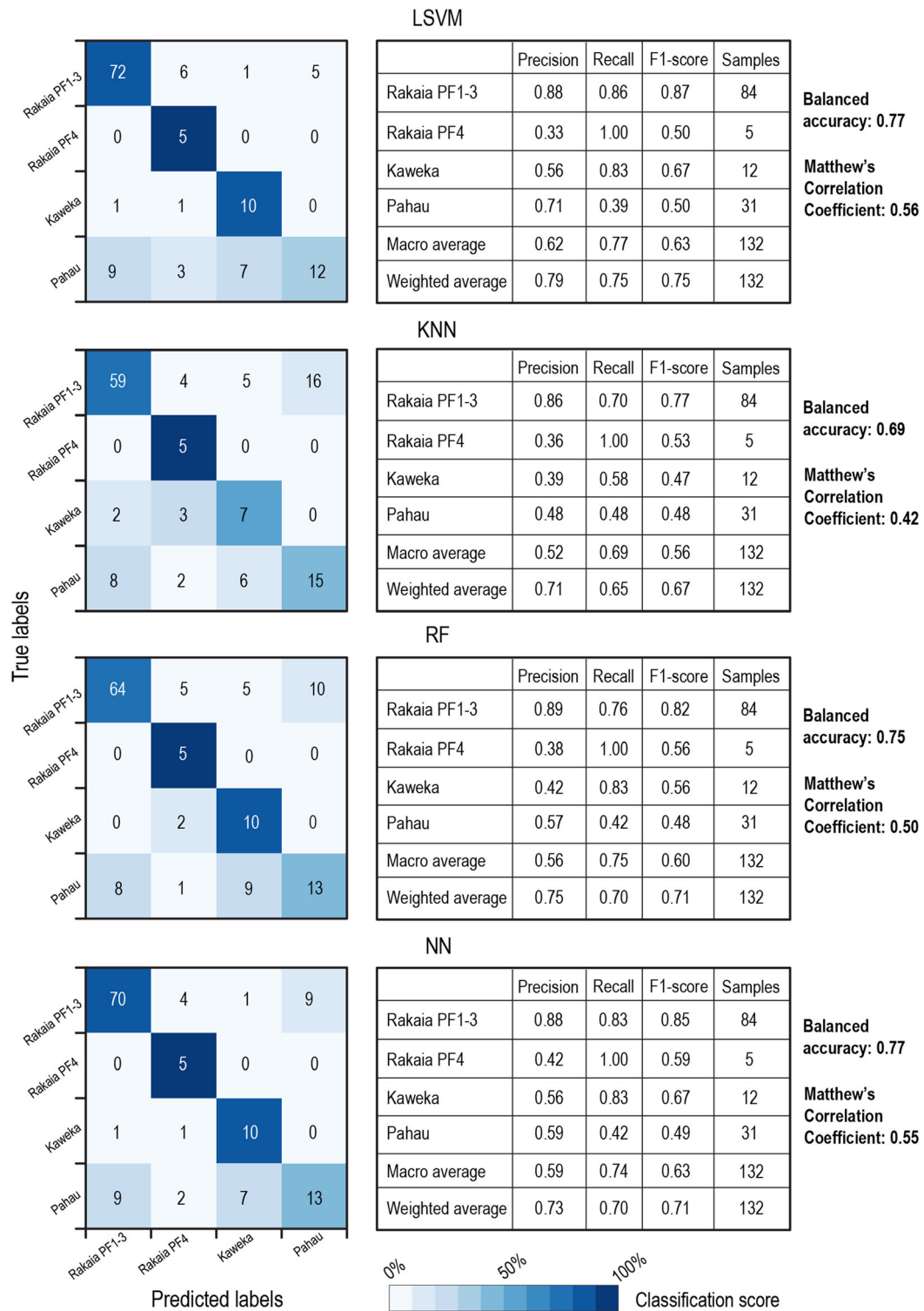


FIGURE 6 | Confusion matrices and classification performance statistics for the four machine learning algorithms used in this study. KNN = k-nearest neighbour, LSVM = linear support vector machine, NN = neural network, RF = random forest.

classified) is Rakaia Terrane PF4, which has a perfect score (1.00) across all four algorithms. However, this group also has the lowest precision values (proportion of samples predicted to be in a certain group that are from that group) of any petrofacies/terrane (0.33–0.42). The Kaweka Terrane also appears to follow a similar pattern to the Rakaia Terrane PF4, although less extreme, with recall values varying from 0.58 to

0.83 and precision values 0.39 to 0.56. The Rakaia Terrane PF1–3 had the highest precision (0.86–0.89) and F1-scores (0.77–0.87), indicating that unknown samples classified into this group have the highest likelihood of being correct. However, there does appear to be a consistent proportion of Rakaia Terrane PF1–3 samples that are being classified as being Pahau Terrane (6–19%), and an even larger proportion

of Pahau Terrane samples that are being classified as Rakaia Terrane PF1–3 (26–29%). This may be related to the Pahau Terrane being derived at least in part from older Rakaia Terrane petrofacies, especially PF1–3 (MacKinnon 1983; Roser and Korsch 1999; Wandres et al. 2005).

To further evaluate the accuracy of our classification, we used an equal voting system where the more algorithms that predict the same petrofacies/terrane, the higher the assigned confidence category. This results in five potential confidence categories: A – all four algorithms predict the same petrofacies/terrane; B – three algorithms predict the same petrofacies/terrane with one dissenting; C – two predict petrofacies/terrane, with each gaining the votes of two algorithms, D – three predict petrofacies/terrane, with one having a majority of two algorithms; and E – all algorithms predict different petrofacies/terrane. For categories A and B, we assigned the predicted petrofacies/terrane, whereas for category C, we assigned both predicted petrofacies/terrane classifications, which are equally shared. For categories D and E, we did not assign a petrofacies/terrane due to the disagreement of the algorithms.

Of the test dataset, 66% samples (87/132) are classified as category A, 20% (26/132) as category B, 13% (17/132) as category C, and 2% (2/132) as category D. However, a small proportion of samples in the test set (20%, 26/132) were predicted as being the incorrect petrofacies/terrane, with 12% (16/132) being category A. This indicates that if there is a complete agreement between all or three of the algorithms it does not always mean that the resulting prediction has a higher likelihood of being correct. Using the equal voting system, the balanced accuracy is 72% using correctly classified category A and B samples. This rises to 80% and a Matthew's Correlation Coefficient of 0.63 when combined with

category C samples (shared predictions), where two algorithms predict the correct petrofacies/terrane and other two algorithms predict another terrane/petrofacies (Figure 7). These metrics are slightly higher than the balanced accuracies of the individual algorithms (69%–77%), indicating that aggregating the algorithms in the equal voting system is still favourable despite some samples being incorrectly classified while simultaneously having a high confidence category.

5 | Discussion

5.1 | Comparisons to Existing TCT Geochemical Classifications

To compare the accuracy of our machine-learning classification with the Th/Sc–SiO₂/Al₂O₃ plot (Roser and Korsch 1999), we used the same external test set that was used to calculate the accuracy of the four machine-learning algorithms and equal voting system and assess where these same samples plot relative to the Rakaia PF1–3, Rakaia PF4, Pahau and volcanogenic terrane (VT) fields. We used a separate python code where the field borders were treated as decision boundaries and tracked the sample IDs. As there is no Kaweka Terrane field on this plot, the 12 Kaweka Terrane samples were not used. The classification statistics and confusion matrix are shown with the equal voting system classification results for direct comparison (Figure 7). The overall performance of this classification is similar to the four individual algorithms and the equal voting system, with a balanced accuracy of 77% and Matthew's Correlation Coefficient of 0.63. Both the Rakaia PF1–3 and Rakaia PF4 have high recall values (0.81 and 0.80, respectively), which are slightly lower than the equal voting system classification (0.88 and 1.00), however the

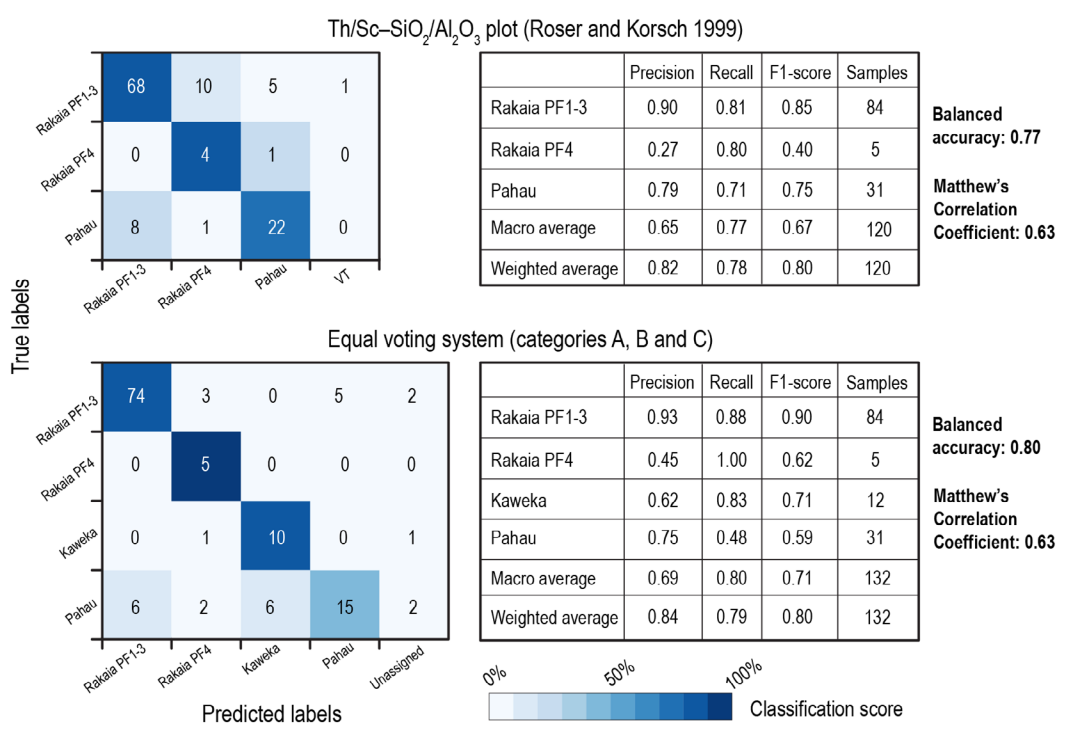


FIGURE 7 | Confusion matrices and classification performance statistics comparing the Th/Sc–SiO₂/Al₂O₃ plot to our equal voting system, where all four algorithms contribute to assigned a petrofacies/terrane to a sample.

Pahau Terrane recall is substantially higher in the Th/Sc–SiO₂/Al₂O₃ classification (0.71 vs. 0.48), indicating that Pahau Terrane samples are more likely to be correctly classified using this classification method.

Interestingly, only 65% of the samples in the external test set have the same classification in both the Th/Sc–SiO₂/Al₂O₃ plot and the equal voting system despite both classifications having a similar accuracy (Kaweka Terrane samples excluded). This lowers further to 56%–63% when comparing each of the four machine-learning algorithms to the Th/Sc–SiO₂/Al₂O₃ plot. Related to this, where both classification appear to be underperforming (e.g., Pahau Terrane samples being classified as Rakaia Terrane PF1–3), almost all of samples that are being misclassified are not misclassified in both the Th/Sc–SiO₂/Al₂O₃ plot and the equal voting system. These differences in petrofacies/terrane classifications between classification methods are likely a result of the greater number of element ratios being used in the machine learning algorithms.

The LDA classification from [Parker \(2022\)](#) had high correct cross-validated reported classification rates (Rakaia PF1–3 = 95.8%, Rakaia PF4 = 95.8%, Pahau = 71%), but it is not clear if the training data used in this study adhered to the assumptions required for LDA to be statistically robust, including multivariate normality of the distribution of variables within classifications and equality of variance-covariance matrices across classifications (e.g., [Boedeker and Kearns 2019](#)). Because of this, and the fact that this classification was also unable to have a Kaweka Terrane classification option, it is recommended that the equal voting system from this study be used in place of the LDA from [Parker \(2022\)](#).

5.2 | Comparisons to Previous Geoscience Machine Learning Studies

These accuracies are comparable to other machine learning studies that train models using geochemical data and the same or similar machine learning algorithms as used here. While some previous comparable studies report high accuracy (>90%) ([Petrelli and Perugini 2016](#); [Petrelli et al. 2017](#); [Ueki et al. 2018](#); [Doucet et al. 2022](#); [Ou et al. 2024](#)), these rates appear to be related to cross-validation using the training dataset, where the algorithms have already seen the data. This can increase accuracy through overfitting, which ultimately results in poorer accuracy on unseen samples ([Tamanna et al. 2025](#)).

Where studies have validated models using unseen data, the classification accuracy is lower. In a study using geochemical data to classify the tectonic environments of oceanic and continental basalts, [Doucet et al. \(2022\)](#) reported high accuracy (recall mean = 88%) using a k-fold cross-validation decision tree, LSVM, and RF algorithms, however, on samples where the tectonic environment labels were removed, the accuracy was reduced to an average of 50% (range = 32%–62%). Similar accuracies are reported in [Dornan et al. \(2020\)](#), who used geochemical data to classify problematic aggregate quarry sources in Ireland using logistic regression and RF algorithms using known unknown samples. Their accuracy for the logistic regression classifier is 67%, while the RF algorithm had an accuracy of 40%.

While these studies have lower rates than our study, a recent paper has reported a 99% classification accuracy using fully connected NN models to classify samples according to the TAS plot ([Tamanna et al. 2025](#)). This is much higher than our accuracies of individual algorithms (69%–77%) but may be a result of the greater existing geochemical variation between the different classification options on a TAS diagram, while in our study, TCT samples form a small elliptical cluster that is confined in the rhyolite and dacite fields on a TAS diagram (Figure S4). Overall, our model performance is comparable to other similar machine learning classifications that utilise geochemical data.

5.3 | Provenance Example Usage

Using the TCT-derived conglomerate clast dataset and the confidence categories defined from the external test dataset, 27 out of 38 (71%) of the samples fall into category A, seven (24%) in category B, and two (5%) in category C (Table 1). There were no samples in categories D or E. This indicates a generally good agreement between the four different classification algorithms and indicates that they are performing well on the independent dataset. As the accuracy of the equal voting system is 80%, approximately eight conglomerate clasts within the dataset can be expected to be incorrectly classified.

The sample locations and predicted clast provenance are shown in Figure 8 with paleogeographic reconstructions from [Seebeck et al. \(2023\)](#) at three time periods corresponding to the stratigraphic age of the samples. The predicted provenance of the majority of samples on both the Pacific and Australian plates can be reasonably interpreted to have been derived from the mapped modern basement extents of the TCT petrofacies/terrane via fluvial/submarine sediment transport (<150 km transport). However, three sandstone clasts from Cardrona Valley (CV) in Otago are predicted to be derived from the Rakaia Terrane PF4. With the current mapped extents of the TCT petrofacies/terrane, at its closest point, the Rakaia Terrane PF4 is still >270 km palaeoshoreline-parallel ([Strogen et al. 2023](#)) to the northeast. Despite paleocurrent directions and the presence of rare lamprophyre clasts from the Alpine Dike Swarm indicating transport from the north and northeast ([McDonnell and Craw 2003](#)), these distances are probably too great to transport cobble-sized clasts from the currently mapped extents of Rakaia Terrane PF4. These clasts could therefore represent misclassifications, as some of the anticipated eight samples were incorrectly classified.

Alternatively, these clasts could have been derived from now-exhumed areas of sandstone closer to the Alpine Fault (now TZIII–IV schist) with an unknown composition that may have been more similar to the Rakaia Terrane PF4. Recent work indicating mid-Cretaceous maximum depositional ages for parts of the Haast Schist in Otago and along the Alpine Fault ([Cooper and Ireland 2015](#); [Cooper and Palin 2018](#); [Mortimer et al. 2024](#)) suggests that there are currently unresolved complexities within the Haast Schist. Another possibility is that these sandstone conglomerate clasts could have been recycled from conglomerate clasts within the TCT itself and/or basal Late Cretaceous units

TABLE 1 | Modified output of classification python code for TCT-derived sandstone conglomerate clast samples showing classification results and confidence category for each algorithm.

Sample ID	Sampling location	Petrofacies/terrane classification				Confidence category	Equal voting system classification
		LSVM	KNN	RF	NN		
UC22314	1	1	5	5	5	B	Pahau
UC22311	1	5	5	1	5	B	Pahau
UC22310	1	5	5	5	5	A	Pahau
UC22308	1	6	6	6	6	A	Kaweka
UC22330	2	6	4	6	6	B	Kaweka
UC22331	2	4	4	4	4	A	Rakaia PF4
UC22334	3	1	1	1	1	C	Rakaia PF1-3/Kaweka
UC22333	3	1	1	1	1	A	Rakaia PF1-3
UC22332	3	1	1	1	1	A	Rakaia PF1-3
UC22337	4	5	5	5	5	A	Pahau
UC22336	4	6	4	4	4	B	Kaweka
UC22317	5	1	1	1	1	A	Rakaia PF1-3
UC22319	5	1	1	1	1	A	Rakaia PF1-3
UC22321	6	6	6	6	6	A	Kaweka
UC22320	6	6	6	6	6	A	Kaweka
UC22323	7	6	6	6	6	A	Kaweka
UC22324	7	6	6	6	6	A	Kaweka
UC22325	7	6	6	6	6	A	Kaweka
UC22329	8	5	5	5	5	A	Pahau
UC22301	9	6	6	6	6	A	Kaweka
UC22299	9	6	4	6	6	A	Kaweka
UC22298	9	4	4	4	4	A	Rakaia PF4
UC22297	9	4	4	4	4	A	Rakaia PF4
UC22305	10	6	6	6	6	A	Kaweka
UC22306	10	6	6	6	6	A	Kaweka
UC22304	10	4	4	4	4	A	Rakaia PF4
UC22305	10	4	4	1	1	C	Rakaia PF1-3/Rakaia PF4
P55265	11	1	1	1	1	A	Rakaia PF1-3
P55267	11	1	1	1	1	A	Rakaia PF1-3
P55291	12	1	4	1	1	B	Rakaia PF1-3
P55292	12	1	5	1	1	B	Rakaia PF1-3
P55293	12	1	4	4	4	B	Rakaia PF4
P55294	12	1	1	1	1	A	Rakaia PF1-3
OU72407	13	4	4	4	4	A	Rakaia PF4
OU72408	13	1	1	1	1	A	Rakaia PF1-3
OU72412	13	1	1	1	1	A	Rakaia PF1-3
OU72413	13	4	5	4	4	B	Rakaia PF4
OU72414	13	4	4	4	4	A	Rakaia PF4

Note: LSVM = linear support vector machine, KNN = k-nearest neighbours, RF = random forest, NN = neural network. 1 = Rakaia Terrane PF1-3, 4 = Rakaia Terrane PF4, 5 = Pahau Terrane, 6 = Kaweka Terrane. Samples with the prefix "UC" are from [Parker \(2022\)](#), those with "P" are from [Mortimer et al. \(2001\)](#), and those with "OU" are from [McDonnell and Craw \(2003\)](#).

directly overlying the TCT (e.g., Broken River Formation, Bluff Sandstone) (e.g., [Crampton 1988](#); [Wandres 2002](#)). While recycling of Late Cretaceous–Cenozoic sand-sized grains stratigraphically

underlying where sandstone conglomerate clasts were sampled has been documented to occur ([McConchie and Lewis 1978](#); [Parker 2022](#)), the thin stratigraphic thicknesses and highly

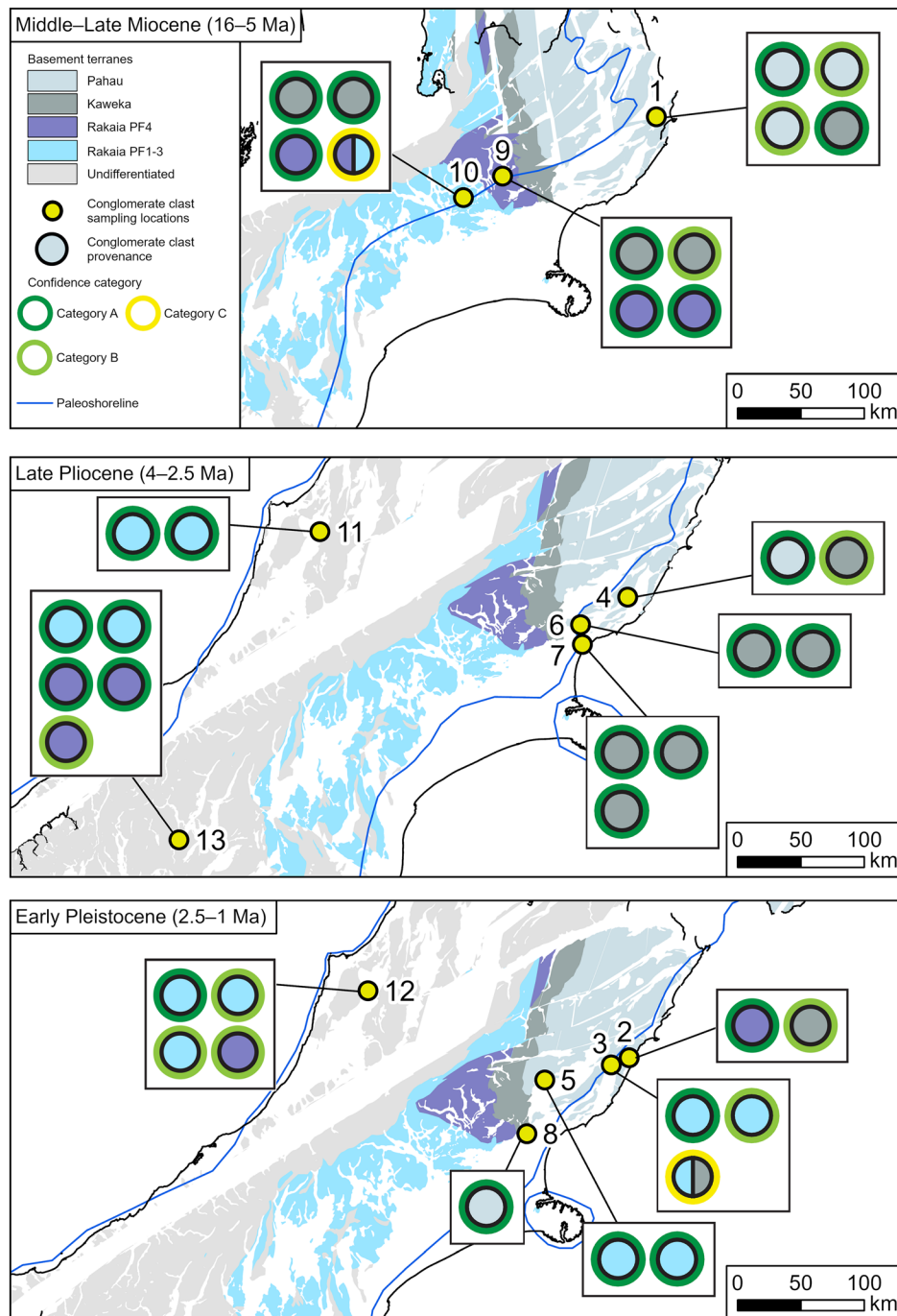


FIGURE 8 | Paleogeographic reconstructions of Te Waipounamu/the South Island at three time intervals showing the position of the terranes/petrofacies of the Torlesse Composite Terrane (TCT), and sampling locations where TCT-derived conglomerate clasts were collected, and conglomerate clast predicted provenance and confidence categories. Paleogeographic reconstructions from [Seebeck et al. \(2023\)](#). Paleoshoreline position from [Parker \(2022\)](#) and [Strogen et al. \(2023\)](#).

localised occurrence of conglomerate units within the TCT and older cover units mean that recycling of cobble-sized clasts is considered extremely unlikely.

In any case, the use of this classification method for conglomerate clast provenance should be treated as one tool in a wider provenance study, with other techniques like paleocurrent directions, clast counts, clast petrography, and detrital zircon U–Pb geochronology helping to validate the provenance

predictions. To truly test the effectiveness of this classification, modern clasts from catchments in the Southern Alps with a known petrofacies/terran affiliation would need to be analysed. Additionally, it could provide an opportunity to select conglomerate clasts of varying weathering intensity to test the effects that weathering has on classification. Weathering intensity could be measured by CIA ([Nesbitt and Young 1984](#)) and PIA indices ([Fedó et al. 1995](#)) or the MFW plot ([Ohta and Arai 2007](#)).

6 | Conclusions

We have developed an updated classification method for South Island TCT petrofacies/terrane that utilises four separate machine learning algorithms (support vector machine, k-nearest neighbour, random forest, and neural network) to classify sandstone samples into one of four petrofacies/terrane (Rakaia Terrane PF1–3, Rakaia Terrane PF4, Kaweka Terrane, and the Pahau Terrane). Previous classifications did not include the Kaweka Terrane as a classification option, as it had not been recognised as being distinct. This has been updated here. Overall balanced accuracies for the four algorithms are 69%–77%, which rises to 80% when an equal voting system is used. These are similar to previous classifications and comparable to other machine learning studies using geochemical data. Because the elements used are generally considered to be immobile, this method could be applied to classify the provenance of weathered conglomerate clasts. Using the available late Cenozoic TCT-derived sandstone conglomerate clast samples with whole-rock trace element geochemistry as a case study, we show how this method can be used to determine sediment provenance. A closer estimate of the accuracy and usefulness/robustness of this technique would ultimately be determined by testing modern clasts from catchments that erode a single terrane/petrofacies of the TCT.

Acknowledgments

The authors thank Kathleen Marsaglia, Barry Roser and an anonymous reviewer for their helpful comments and suggestions that significantly improved the manuscript during the peer review process.

Open access publishing facilitated by University of Canterbury, as part of the Wiley - University of Canterbury agreement via the Council of Australasian University Librarians.

Funding

This study was supported by the Strategic Science Investment Fund (SSIF) from the Ministry of Business, Innovation and Employment (MBIE) to GNS Science.

Data Availability Statement

The data that supports the findings of this study are available in the supplementary material of this article.

References

- Adams, C. J., H. J. Campbell, and W. J. Griffin. 2008. "Age and Provenance of Basement Rocks of the Chatham Islands: An Outpost of Zealandia." *New Zealand Journal of Geology and Geophysics* 51, no. 3: 245–259.
- Adams, C. J., H. J. Campbell, N. Mortimer, and W. L. Griffin. 2017. "Perspectives on Cretaceous Gondwana Break-up from Detrital Zircon Provenance of Southern Zealandia Sandstones." *Geological Magazine* 154, no. 4: 661–682.
- Adams, C. J., N. Mortimer, H. J. Campbell, and W. L. Griffin. 2011. "Recognition of the Kaweka Terrane in Northern South Island, New Zealand: Preliminary Evidence from Rb–Sr Metamorphic and U–Pb Detrital Zircon Ages." *New Zealand Journal of Geology and Geophysics* 54, no. 3: 291–309.
- Adams, C. J., N. Mortimer, H. J. Campbell, and W. L. Griffin. 2013. "The Mid-Cretaceous Transition from Basement to Cover Within Sedimentary Rocks in Eastern New Zealand: Evidence from Detrital Zircon Age Patterns." *Geological Magazine* 150, no. 3: 455–478.
- Aitchison, J. 1986. *The Statistical Analysis of Compositional Data*. Chapman and Hall.
- Andrews, P. B., I. G. Speden, and J. D. Bradshaw. 1976. "Lithological and Paleontological Content of the Carboniferous–Jurassic Canterbury Suite, South Island New Zealand." *New Zealand Journal of Geology and Geophysics* 19, no. 6: 791–819.
- Begg, J. G., and C. Mazengarb. 1996. *Geology of the Wellington Area, Scale 1: 50 000. Institute of Geological and Nuclear Sciences Geological Map 22. 1 Sheet + 128 Lower Hutt*. Institute of Geological & Nuclear Sciences Limited.
- Beggs, J. M. 1980. "Sedimentology and Paleogeography of Some Kaihikuan Torlesse Rocks in Mid Canterbury." *New Zealand Journal of Geology and Geophysics* 23, no. 4: 439–445.
- Boedeker, P., and N. T. Kearns. 2019. "Linear Discriminant Analysis for Prediction of Group Membership: A User-Friendly Primer." *Advances in Methods and Practices in Psychological Science* 2, no. 3: 250–263.
- Botsford, J. W. 1983. *The Esk Head Melange in the Esk Head/Okuku Area, North Canterbury [Unpublished MSc Thesis]*. University of Canterbury. 249.
- Bradshaw, J. D. 1972. "Stratigraphy and Structure of the Torlesse Supergroup (Triassic–Jurassic) in the Foothills of the Southern Alps near Hawarden (S60–61) Canterbury." *New Zealand Journal of Geology and Geophysics* 15, no. 1: 71–87.
- Bradshaw, J. D. 1973. "Allochthonous Mesozoic Fossil Localities in Melange Within the Torlesse Rocks of North Canterbury." *Journal of the Royal Society of New Zealand* 3, no. 2: 161–167.
- Browne, G. H., and T. R. Naish. 2003. "Facies Development and Sequence Architecture of a Late Quaternary Fluvial-Marine Transition, Canterbury Plains and Shelf, New Zealand: Implications for Forced Regressive Deposits." *Sedimentary Geology* 158, no. 1–2: 57–86.
- Campbell, J. D., and G. Warren. 1965. "Fossil Localities of the Torlesse Group in the South Island. Transactions of the Royal Society of New Zealand." *Geology* 3, no. 8: 99–137.
- Cooper, A. F., and T. R. Ireland. 2015. "The Pounamu Terrane, a New Cretaceous Exotic Terrane Within the Alpine Schist, New Zealand; Tectonically Emplaced, Deformed and Metamorphosed during Collision of the LIP Hikurangi Plateau with Zealandia." *Gondwana Research* 27, no. 3: 1255–1269.
- Cooper, A. F., and J. M. Palin. 2018. "Two-Sided Accretion and Polyphase Metamorphism in the Haast Schist Belt, New Zealand: Constraints from Detrital Zircon Geochronology." *GSA Bulletin* 130, no. 9–10: 1501–1518.
- Crampton, J. S. 1988. "Stratigraphy and Structure of the Monkey Face Area, Marlborough, New Zealand, with Special Reference to Shallow Marine Cretaceous Strata." *New Zealand Journal of Geology and Geophysics* 31, no. 4: 447–470.
- Crampton, J. S., and C. A. Landis. 1988. "Petrography of Jurassic–Cretaceous Sedimentary Rocks of the Monkey Face Area, Southern Marlborough." *New Zealand Geological Survey Record* 35: 77–85.
- Cutten, H. N. C., R. J. Korsch, and B. P. Roser. 2006. "Using Geochemical Fingerprinting to Determine Transpressive Fault Movement History: Application to the New Zealand Alpine Fault." *Tectonics* 25, no. 4: TC4014.
- Di Martino, A., G. Carlini, G. Castellani, D. Remondini, and A. Amorosi. 2023. "Sediment Core Analysis Using Artificial Intelligence." *Scientific Reports* 13, no. 1: 20409.
- Dickinson, W. R. 1971. "Detrital Modes of New Zealand Graywackes." *Sedimentary Geology* 5, no. 1: 37–56.

- Dickinson, W. R., and E. I. Rich. 1972. "Petrologic Intervals and Petrofacies in the Great Valley Sequence, Sacramento Valley, California." *Geological Society of America Bulletin* 83, no. 10: 3007–3024.
- Dornan, T., G. O'Sullivan, N. O'Riain, E. Stueeken, and R. Goodhue. 2020. "The Application of Machine Learning Methods to Aggregate Geochemistry Predicts Quarry Source Location: An Example from Ireland." *Computers & Geosciences* 140: 104495.
- Doucet, L. S., M. G. Tetley, Z.-X. Li, Y. Liu, and H. Gamaledien. 2022. "Geochemical Fingerprinting of Continental and Oceanic Basalts: A Machine Learning Approach." *Earth-Science Reviews* 233: 104192.
- Fedo, C. M., H. Wayne Nesbitt, and G. M. Young. 1995. "Unraveling the Effects of Potassium Metasomatism in Sedimentary Rocks and Paleosols, with Implications for Paleoweathering Conditions and Provenance." *Geology* 23, no. 10: 921–924.
- Field, B. D., and G. H. Browne. 1989. "Cretaceous and Cenozoic Sedimentary Basins and Geological Evolution of the Canterbury Region, South Island, New Zealand." *New Zealand Geological Survey Basin Studies* 2.
- Fleming, S. W., J. R. Watson, A. Ellenson, A. J. Cannon, and V. C. Vesselinov. 2021. "Machine Learning in Earth and Environmental Science Requires Education and Research Policy Reforms." *Nature Geoscience* 14, no. 12: 878–880.
- Folk, R. L., P. B. Andrews, and D. W. Lewis. 1970. "Detrital Sedimentary Rock Classification and Nomenclature for use in New Zealand." *New Zealand Journal of Geology and Geophysics* 13, no. 4: 937–968.
- Forsyth, P. J., D. J. A. Barrell, and R. Jongens. 2008. "Geology of the Christchurch Area," In *Institute of Geological and Nuclear Sciences 1:250,000 geological map 16, 1 sheet + 67 p. Lower Hutt, New Zealand*. Institute of Geological & Nuclear Sciences Limited.
- Gaina, C., D. R. Müller, J. Y. Royer, J. Stock, J. Hardebeck, and P. Symonds. 1998. "The Tectonic History of the Tasman Sea: A Puzzle with 13 pieces." *Journal of Geophysical Research: Solid Earth* 103, no. B6: 12413–12433.
- Gooley, J. T., and N. M. Nieminski. 2022. "Detrital Zircon Provenance of the Cretaceous–Neogene East Coast Basin Reveals Changing Tectonic Conditions and Drainage Reorganisation along the Pacific Margin of Zealandia." *Geosphere* 18, no. 2: 616–646.
- Hassan, A., S. Chan, M. Mahmoud, M. S. Aljawad, J. Humphrey, and A. Abdulraheem. 2022. "Artificial Intelligence-Based Model of Mineralogical Brittleness Index Based on Rock Elemental Compositions." *Arabian Journal for Science and Engineering* 47, no. 9: 11745–11761.
- Heron, D. W. 2020. "Geological Map of New Zealand 1: 250 000." In *GNS Science Geological Map 1. Lower Hutt, GNS Science, (3rd)*.
- Hicks, D. M. 1981. "Deep-sea fan sediments in the Torlesse zone, Lake Ohau, South Canterbury, New Zealand." *New Zealand Journal of Geology and Geophysics* 24, no. 2: 209–230.
- Lachowycz, S. 2024. "Utility of Artificial Intelligence in Geoscience." *Nature Geoscience* 17, no. 10: 953–955.
- Landis, C. A., and D. G. Bishop. 1972. "Plate Tectonics and Regional Stratigraphic-Metamorphic Relations in the Southern Part of the New Zealand Geosyncline." *Geological Society of America Bulletin* 83, no. 8: 2267–2284.
- Lui, T. C., D. D. Gregory, M. Anderson, W.-S. Lee, and S. A. Cowling. 2022. "Applying Machine Learning Methods to Predict Geology Using Soil Sample Geochemistry." *Applied Computing and Geosciences* 16: 100094.
- MacKinnon, T. C. 1980. *Sedimentologic, Petrographic, and Tectonic Aspects of Torlesse and Related Rocks: South Island, New Zealand [Unpublished PhD Thesis]*. University of Otago.
- MacKinnon, T. C. 1983. "Origin of the Torlesse Terrane and Coeval Rocks, South Island New Zealand." *Geological Society of America Bulletin* 94, no. 8: 967–985.
- Mandal, P. P., and R. Rezaee. 2019. "Facies Classification with Different Machine Learning Algorithm—An Efficient Artificial Intelligence Technique for Improved Classification." *ASEG Extended Abstracts* 2019, no. 1: 1–6.
- McConchie, D. M., and D. W. Lewis. 1978. "Authigenic, Perigenic, and Allogenic Glauconites from the Castle Hill Basin, North Canterbury." *New Zealand. New Zealand Journal of Geology and Geophysics* 21, no. 2: 199–214.
- McDonnell, M., and D. Craw. 2003. "Stratigraphy and Provenance of Pliocene Greywacke-Bearing Conglomerate, Cardrona Valley, Otago." *New Zealand. New Zealand Journal of Geology and Geophysics* 46, no. 3: 425–436.
- Mortimer, N. 1994. "Origin of the Torlesse Terrane and Coeval Rocks, North Island, New Zealand." *International Geology Review* 36, no. 10: 891–910.
- Mortimer, N. 2004. "New Zealand's Geological Foundations." *Gondwana Research* 7, no. 1: 261–272.
- Mortimer, N. 2025. "Tectonics, Geology and Origins of Te Riu-a-Māui/ Zealandia." *New Zealand Journal of Geology and Geophysics* 68, no. 4: 531–567.
- Mortimer, N., J. Lee, and D. F. Stockli. 2024. "Terrane and Core Complex Architecture of the Otago Schist in the Dunstan and Cairnmuir Mountains, New Zealand, from U-Pb and (U-Th)/He Zircon Dating." *New Zealand Journal of Geology and Geophysics* 67, no. 2: 195–208.
- Mortimer, N., M. S. Rattenbury, P. R. King, et al. 2014. "High-Level Stratigraphic Scheme for New Zealand Rocks." *New Zealand Journal of Geology and Geophysics* 57, no. 4: 402–419.
- Mortimer, N., R. Sutherland, and S. Nathan. 2001. "Torlesse Greywacke and Haast Schist Source for Pliocene Conglomerates near Reefton New Zealand." *New Zealand Journal of Geology and Geophysics* 44, no. 1: 105–111.
- Nesbitt, H. W., and G. M. Young. 1984. "Prediction of Some Weathering Trends of Plutonic and Volcanic Rocks Based on Thermodynamic and Kinetic Considerations." *Geochimica et Cosmochimica Acta* 48, no. 7: 1523–1534.
- Ohta, T., and H. Arai. 2007. "Statistical Empirical Index of Chemical Weathering in Igneous Rocks: A New Tool for Evaluating the Degree of Weathering." *Chemical Geology* 240, no. 3–4: 280–297.
- Orlowski, R. J. 2001. *A Sedimentological Profile of the Pahau Terrane Type Locality [Unpublished MSc Thesis]*. University of Canterbury.
- Ou, Q., D. Zheng, A. R. L. Nichols, et al. 2024. "Detachment of the Subducting Slab in Ancient Subduction Zones Discerned Using Machine Learning." *Journal of Geophysical Research: Machine Learning and Computation* 1, no. 4: e2024JH000400.
- Parker, M. O. 2022. *Provenance of Middle Miocene–early Pleistocene Conglomerates in the Northern Canterbury Basin, Northeastern South Island, New Zealand [Unpublished MSc Thesis]*. University of Canterbury.
- Petrelli, M., R. Bizzarri, D. Morgavi, A. Baldanza, and D. Perugini. 2017. "Combining Machine Learning Techniques, Microanalyses and Large Geochemical Datasets for Tephrochronological Studies in Complex Volcanic Areas: New Age Constraints for the Pleistocene Magmatism of Central Italy." *Quaternary Geochronology* 40: 33–44.
- Petrelli, M., and D. Perugini. 2016. "Solving Petrological Problems through Machine Learning: The Study Case of Tectonic Discrimination Using Geochemical and Isotopic Data." *Contributions to Mineralogy and Petrology* 171: 1–15.

- Rattenbury, M. S., D. B. Townsend, and M. R. Johnston. 2006. "Geology of the Kaikoura Area," In *Institute of Geological & Nuclear Sciences 1: 250 000 Geological Map 13. 1 sheet + 70 p. Lower Hutt, New Zealand*. Institute of Geological & Nuclear Sciences Limited.
- Riefstahl, F., K. Gohl, B. Davy, et al. 2020. "Cretaceous Intracontinental Rifting at the Southern Chatham Rise Margin and Initialisation of Seafloor Spreading between Zealandia and Antarctica." *Tectonophysics* 776: 228298.
- Roser, B., R. Cooper, S. Nathan, and A. Tulloch. 1996. "Reconnaissance sandstone geochemistry, provenance, and tectonic setting of the lower Paleozoic terranes of the West Coast and Nelson, New Zealand." *New Zealand Journal of Geology and Geophysics* 39, no. 1: 1–16.
- Roser, B. P., D. S. Coombs, R. J. Korsch, and J. D. Campbell. 2002. "Whole-rock geochemical variations and evolution of the arc-derived Murihiku Terrane, New Zealand." *Geological Magazine* 139, no. 6: 665–685.
- Roser, B. P., R. H. Grapes, and R. J. Korsch. 1995. "XRF Analyses of Sandstones and Argillites from the Torlesse Terrane, New Zealand." Victoria University of Wellington, Research School of Earth Sciences, Geology Board of Studies Publication.
- Roser, B. P., R. H. Grapes, and K. Palmer. 1993. "Petrofacies and Subterranean Affiliation of Torlesse Sediments, Southern North Island, New Zealand." *Geoscience Society of New Zealand Miscellaneous Publication 79A Geoscience Society of New Zealand*.
- Roser, B. P., and R. J. Korsch. 1986. "Determination of Tectonic Setting of Sandstone-Mudstone Suites Using SiO₂ Content and K₂O/Na₂O Ratio." *The Journal of Geology* 94, no. 5: 635–650.
- Roser, B. P., and R. J. Korsch. 1988. "Provenance Signatures of Sandstone-Mudstone Suites Determined Using Discriminant Function Analysis of Major-Element Data." *Chemical Geology* 67, no. 1-2: 119–139.
- Roser, B. P., and R. J. Korsch. 1999. "Geochemical Characterization, Evolution and Source of a Mesozoic Accretionary Wedge: The Torlesse Terrane, New Zealand." *Geological Magazine* 136, no. 5: 493–512.
- Rotzien, J. R., G. H. Browne, and P. R. King. 2018. "Geochemical, Petrographic, and Uranium-lead Geochronological Evidence for Multisourced Polycyclic Provenance of Deep-Water Strata in a Hybrid Tectonic Setting: The Upper Miocene Upper Mount Messenger Formation, Taranaki Basin, New Zealand." *AAPG Bulletin* 102, no. 9: 1763–1802.
- Sagar, M. W., K. E. Higgs, D. P. Strogen, K. J. Bland, and G. H. Browne. 2024. "Sediment Provenance in the Murchison and Maruia Basins, Aotearoa/New Zealand: A Record of Neogene Strike-Slip Displacement, Convergence, and Basement Exhumation along the Australian-Pacific Plate Boundary." *New Zealand Journal of Geology and Geophysics* 67, no. 1: 45–83.
- Seebeck, H., D. P. Strogen, A. Nicol, B. R. Hines, and K. J. Bland. 2023. "A Tectonic Reconstruction Model for Aotearoa-New Zealand from the Mid-Late Cretaceous to the Present Day." *New Zealand Journal of Geology and Geophysics* 67, no. 4: 527–550.
- Silberling, N. J., K. M. Nichols, J. D. Bradshaw, and C. D. Blome. 1988. "Limestone and chert in tectonic blocks from the Esk Head subterranean, South Island, New Zealand." *Geological Society of America Bulletin* 100, no. 8: 1213–1223.
- Smale, D. 1978. "The composition of a Torlesse conglomerate—Ethelton, North Canterbury." *New Zealand Journal of Geology and Geophysics* 21, no. 6: 699–711.
- Smale, D. 1990. "Distribution and Provenance of Heavy Minerals in the South Island: A Review." *New Zealand Journal of Geology and Geophysics* 33, no. 4: 557–571.
- Strogen, D. P., H. Seebeck, B. R. Hines, K. J. Bland, and J. S. Crampton. 2023. "Palaeogeographic Evolution of Zealandia: Mid-Cretaceous to Present." *New Zealand Journal of Geology and Geophysics* 66, no. 3: 528–557.
- Strong, D. T., R. E. Turnbull, S. Haubrock, and N. Mortimer. 2016. "Petlab: New Zealand's National Rock Catalogue and Geoanalytical Database." *New Zealand Journal of Geology and Geophysics* 59, no. 3: 475–481.
- Tamanna Hezel, D. C., N. Srivastava, and J. Faber. 2025. "Using Machine Learning for Automatic Rock Classification." *American Mineralogist* 111, no. 4: 523–533.
- Taylor, S. R., and S. M. McLennan. 1985. *The Continental Crust: Its Composition and Evolution*. Blackwell Scientific.
- Turnbull, I. M., N. Mortimer, and D. Craw. 2001. "Textural Zones in the Haast Schist—a Reappraisal." *New Zealand Journal of Geology and Geophysics* 44, no. 1: 171–183.
- Ueki, K., H. Hino, and T. Kuwatani. 2018. "Geochemical Discrimination and Characteristics of Magmatic Tectonic Settings: A Machine-Learning-Based Approach." *Geochemistry, Geophysics, Geosystems* 19, no. 4: 1327–1347.
- Wandres, A. M. 2002. *Provenance Study of the Torlesse Terranes and Implications for the Origin of the Continental Crust of Eastern New Zealand [Unpublished PhD Thesis]*. University of Canterbury.
- Wandres, A. M., J. D. Bradshaw, and T. Ireland. 2005. "The Paleozoic-Mesozoic Recycling of the Rakaia Terrane, South Island, New Zealand: Sandstone Clast and Sandstone Petrology, Geochemistry, and Geochronology." *New Zealand Journal of Geology and Geophysics* 48, no. 2: 229–245.
- Wood, D. A. 2021. "Enhancing Lithofacies Machine Learning Predictions with Gamma-Ray Attributes for Boreholes with Limited Diversity of Recorded Well Logs." *Artificial Intelligence in Geosciences* 2: 148–164.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Mafic-felsic-weathered (MFW) ternary plot (Ohta and Arai 2007) comparing TCT petrofacies/terrane and late Cenozoic sandstone conglomerate clasts. TCT data from Roser et al. (1995), Wandres (2002), the Petlab database (Strong et al. 2016) and this study. Sandstone conglomerate clast data from Mortimer et al. (2001), McDonnell and Craw (2003) and Parker (2022). **Figure S2:** Chemical index of alteration (Nesbitt and Young 1984) (left) and plagioclase index of alteration (Fedo et al. 1995) plots comparing TCT petrofacies/terrane and late Cenozoic sandstone conglomerate clasts. Data sources are identical to Figure S1. **Figure S3:** Bar graph showing the feature importances of the log-transformed ratios for the random forest algorithm. Feature importances add to 1. **Figure S4:** Total alkali-silica (TAS) classification plot showing clustering of TCT data. Data sources are identical to Figure S1.